Reliability Aspects in Resistively Switching Valence Change Memory Cells

Von der Fakultät für Elektrotechnik und Informationstechnik der
Rheinisch-Westfälischen Technischen Hochschule Aachen zur
Erlangung des akademischen Grades eines Doktors der
Ingenieurwissenschaften genehmigte Dissertation
vorgelegt von

Stefan Wiefels, M.Sc.
aus Grevenbroich

Berichter: Univ.-Prof. Dr.-Ing. Rainer Waser
Univ.-Prof. Dr.-Ing. Christoph Jungemann

Tag der mündlichen Prüfung: 06.04.2021

Abstract

For over 50 years, Moore's law functioned as road map for advancements in the semiconductor industry [1]. Soon, the predicted exponential increase in the number of devices per microchip will reach physical limitations [2]. In order to overcome these limitations, redox-based resistive switching random access memory (ReRAM) is discussed as promising candidate for future memory applications. Recently, also a potential application of ReRAM in neuro-inspired architectures is gaining a lot of attention. Among other approaches, valence change based memory (VCM) is studied intensively.

Regardless of an application as classical memory or as neuronal network component, the reliability of ReRAM devices is the key attribute for industrial adaption. This dissertation addresses the three main components of the reliability VCM ReRAM devices, being variability, retention and endurance. Here, VCM ReRAM cells based on ZrO_2 fabricated under laboratory conditions are characterized as well as industrial devices based on HfO_2 as switching oxide. Throughout this work, a focus on large arrays instead of single cells is emphasized. The evaluation and interpretation is focused on the internal statistics rather than on the behavior of individual devices.

The variability of VCM ReRAM remains one of the largest challenges for their large scale adaption in industrial applications. Whereas the stochastic nature of the switching process can be significantly reduced by appropriate programming algorithms, random fluctuations occur also between read operations. This read to read (R2R) variability is identified as key challenge in the short term stability of VCM ReRAM. It determines the intrinsic statistics of large memory arrays and effectively limits the read window between the low resistive (LRS) and high resistive state (HRS). The random R2R fluctuations are attributed to random jumps of the conduction supporting oxygen vacancies. In the HRS, these jumps lead to a log-normal read current distribution. Via an empirical model as well as kinetic Monte Carlo (KMC) methods, the most likely origin of these statistics is found to be tunneling across a normally distributed gap in HRS. Here, the exponential dependence of the read current on the tunneling gap results in the observed log-normal statistics.

Investigating the long term stability or retention, the R2R variability remains a key characteristic of the investigated devices. The most critical aspect of the long term degradation of a programmed state is found to be a broadening of the whole distribution, i.e. increasing variability. The trend of the degradation is fitted by an empirical tunneling model which allows for extrapolation of data measured at higher temperatures towards the target retention time at lower operating temperature.

Additionally, a statistical model based on the work of Abbaspour et al. [3–6] is developed which explains the observed degradation by diffusion of oxygen vacancies from a confined filament region towards the active electrode.

Finally, an algorithm is developed which increases the number of possible switching cycles, also referred to as endurance, of a device. It dynamically adjusts the programming parameters to ensure reliable switching. Since the frequency of applied adjustments determines the speed of the experiment, the algorithm dynamically adjusts this frequency to the tested cell. It therefore increases the measurement speed if a cell requires less adjustments. The algorithm is used to determine the maximum endurance for different material combinations. Thus, it is demonstrated that ohmic electrode metals with lower oxygen chemical potential ensure higher endurance which verifies the theoretical findings of Guo et al. [7].
All in all, this dissertation proposes to evaluate the reliability of VCM ReRAM for its intrinsic statistics rather than tracing single cells.

Kurzfassung

Das Moore'sche Gesetz dient seit über 50 Jahren als Fahrplan für die Halbleiterindustrie [1]. In naher Zukunft wird der vorhergesagte exponentielle Anstieg der Anzahl Bauelemente pro Mikrochip an seine physikalischen Grenzen stoßen [2]. Als möglicher Ausweg wird *redox based resistive switching random access memory* (ReRAM) für zukünftige Speicheranwendungen diskutiert. Neben der Anwendung als klassisches Speichermedium, wird ReRAM mittlerweile auch als mögliche Komponente in neuro-inspirierten Architekturen gesehen. Neben anderen Ansätzen gilt *valence change based memory* (VCM) als eine der vielversprechendsten ReRAM Typen. Unabhängig davon ob ReRAM als klassischer Speicher oder als Komponente eines neuronalen Netzes zum Einsatz kommen soll, ist deren Zuverlässigkeit (*reliability*) von großer Bedeutung. Die vorliegende Arbeit befasst sich mit den drei Hauptaspekten der *reliability* von VCM ReRAM. Darunter fallen *variability*, *retention* und *endurance*. Untersucht werden sowohl unter Laborbedingungen hergestellte VCM ReRAM Zellen auf Basis von ZrO_2, als auch industrielle Zellen auf Basis von HfO_2. Der Schwerpunkt der Arbeit liegt dabei auf der Untersuchung von größeren Arrays anstelle von Einzelzellen. Dementsprechend sind Auswertung und Interpretation auf die interne Statistik ausgerichtet.
Die *variability* von VCM ReRAM ist nach wie vor eine der größten Herausforderungen für ihre industrielle Adaption. Während die stochastische Natur des Schaltprozesses durch geeignete Programmieralgorithmen deutlich reduziert werden kann, treten zufällige Schwankungen auch zwischen den Lesevorgängen auf. Diese *Read-to-Read* (R2R) *Variability* kann als eine der größten Herausforderungen genannt werden. Sie bestimmt die intrinsische Statistik von großen Speicher-Arrays und begrenzt maßgeblich das Lesefenster zwischen dem niederohmigen (LRS) und dem hochohmigen Zustand (HRS). In VCM ReRAM Systemen wird die elektrische Leitfähigkeit maßgeblich durch Sauerstoffleerstellen bestimmt. Die zufälligen R2R Fluktuationen werden in dieser Arbeit auf zufällige Sprünge der Sauerstoffleerstellen zurückgeführt. Im HRS führen diese Sprünge zu einer log-normalen Lesestromverteilung. Mit Hilfe eines empirischen Modells sowie von KMC-Methoden wird Tunneln über ein normalverteiltes *gap* als wahrscheinlichste Ursache der Log-Normal-Statistik im HRS identifiziert. Hier führt die exponentielle Abhängigkeit des Lesestroms vom Tunnel-*gap* zu der beobachteten Log-Normal-Verteilung.
Bei der Untersuchung der Langzeitstabilität (*retention*) ist die beschriebene R2R-Variabilität ebenfalls von entscheidender Bedeutung. Die Degradation eines programmierten Zustands äußert sich vor allem in einer Verbreiterung der gesamten Verteilung, d.h. einer zunehmenden Variabilität. Der Trend der Degradation wird in dieser Arbeit durch ein empirisches Tunnelmodell gefittet. Die bei erhöhter Temperatur experimentell bestimmte Degradation kann mithilfe dieses Modells extrapoliert werden, um die *retention* bei der angestrebten Anwendungstemperatur zu ermitteln. Zusätzlich enthält die Arbeit ein statistisches Modell auf der Grundlage der Arbeiten

von Abbaspour et al. [3–6], das die beobachtete Degradation erklärt. Hier führt die Diffusion von Sauerstoffleerstellen aus einem eingegrenzten Filament-Bereich in Richtung der aktiven Elektrode zu einem Anstieg des mittleren HRS Lesestroms und zur Verbreiterung der Verteilung.
Zuletzt wird ein Algorithmus vorgestellt, der die Anzahl der möglichen Schaltzyklen (*endurance*) einer Zelle erhöht. Dieser passt die SET und RESET Spannungen dynamisch an die jeweilige Zelle an, um zuverlässiges Schalten zu garantieren. Da die Häufigkeit solcher Anpassungen die Laufzeit eines solchen Experiments vorgibt, passt der Algorithmus diese Frequenz dynamisch an die getestete Zelle an. Die Messzeit wird also verkürzt, wenn eine Zelle stabil schaltet und weniger Spannungs-Anpassungen erfordert. Der Algorithmus wird außerdem verwendet, um die maximale *endurance* für verschiedene Materialkombinationen zu bestimmen. Durch Variation des Metalls der ohmschen Elektrode wird deutlich, dass Metalle mit niedrigerem chemischen Sauerstoffpotential eine höhere *endurance* ermöglichen. Dies bestätigt die theoretischen Ergebnisse von Guo et al. [7].
Insgesamt liegt der Schwerpunkt der vorliegenden Arbeit auf der statistischen Analyse der Zuverlässigkeit (*reliability*) von VCM ReRAM. Es sollte deutlich werden, dass die Betrachtung der intrinsischen Statistik wesentlich mehr Potenzial bietet, als die Analyse einzelner Zellen.

Acknowledgements

This thesis was written during my time at Infineon Technologies Dresden GmbH, as well as at the Institut für Werkstoffe der Elektrotechnik II (IWE II) at the RWTH Aachen. The work would not have been possible without the help of a lot of people, whom I would like thank at this point.
First of all, I would like to express my gratitude to Prof. Dr. Rainer Waser for the opportunity to do research on the promising field of resistive switching memory and for examining my thesis. Further, I would like to thank Prof. Dr. Christoph Jungemann, who kindly agreed to be the co-examiner of my work.
I am deeply grateful to Dr. Ulrich Böttger for supervising my work and his continuous support. I truly appreciate the many advices, discussions and impulses which helped me pursue my own ideas.
Further, I would like to thank Dr. Stephan Menzel and Dr. Dirk J. Wouters for co-supervising my work, along with extensive input and many fruitful discussions.
I am truly thankful to all the people I worked with at Infineon Technologies for the comprehensive input, experimental data and insights into current questions in industrial non-volatile memory development. In particular, I would like to thank Dr. Karl Hofmann, Dr. Robert Strenz, Dr. Robert Allinger and Dr. Georg Tempel for countless discussions and advices.
I would like to acknowledge Dr. Andreas Kindsmüller for his extensive support regarding device fabrication and for kindly providing the process parameters of his ZrO_2 layers. Further, I very much appreciate the support by Dr. Camilla La Torre and countless discussions during our industry project.
For many fruitful collaborations and the great support regarding modeling and simulation, I would like to thank Christopher Bengel and Carsten Funck.
Special thanks go to Dr. Michael Lübben, who also supervised my Master´s thesis, for his support, encouragement and countless advices and discussions.
Further, I owe many thanks to my student assistants Nils Kopperberg, Michael Hüttemann, Yvonne Wolff, Sergej Liberda and Aaron Verjovsky. Without their extensive input, this thesis would not have been possible.
For all discussions and the great atmosphere, I want to thank my office mates Moritz von Witzleben, Johannes Mohr and Dr. Alexander Schönhals.
Also, I very much appreciate the administrative and technical support by Martina Heins, Udo Evertz, Thomas Pössinger and Dagmar Leisten. Further, I would like to thank Daliborka Erdogilia, Katarina Utens and Petra Grewe for their great support in device fabrication and characterization.
To all colleagues at IWE II in Aachen and the PGI7 in Jülich, I am deeply grateful for countless discussions, an outstanding atmosphere and in general, great four years!
Last but not least, I would like to thank Stefanie Meinken, my family Peter and Ursula Wiefels, Mandy Knetsch and my friends for their constant support, whenever needed.

Contents

1 Introduction

Since the first transistor based computer was introduced in 1965 [8], the demand in memory capacity increased rapidly. Especially the success story of mobile devices (smartphones) starting in the mid-2000s lead to an exploding demand in cheap and efficient memory devices [9]. As predicted by Gordon E. Moore already in 1965 [1], the semiconductor industry managed to increase the number of integrated circuits per microchip exponentially to this day. However, the geometrical down-scaling of conventional, transistor based architectures decelerates as Moore´s law is reaching physical limitations [2].
Therefore, alternative memory concepts gained a lot of attention. Besides magnetoresistive random access memory (MRAM) and phase change memory (PCM), redox based resistive switching random access memory (ReRAM) is considered one of the most promising alternatives for future non-volatile memory applications [10, 11]. Besides their simple two-terminal architecture, ReRAM is preferred due to its non-volatility, fast operation, low power consumption, high endurance and retention and good scalability [9, 12–17].
Historically, resistive switching is studied since the early 1960s [11, 18–20]. In the 1980s the research stagnated due to the advantages of transistor based memories like dynamic random access memory (DRAM) or flash. Limitations in the contemporary analytic tools are believed to be another reason for the stagnating interest in resistive switching memories [10]. In the late 1990s the research was revived by the studies of Asamitsu et al., Kozicki et al., and Beck et al. [10, 12, 13, 21–24]. Starting in the early 2000s, the research interest of the industry increased steadily [9]. To this day, several companies like *Samsung, IBM, Micron, Adesto* and *Panasonic* presented memory concepts and technologies based on ReRAM devices [25–31]. Recently, ReRAM gains increasing attention as potential component in future neuromorphic computing architectures [32–35].
Despite the excellent characteristics of ReRAM devices, their reliability is still intensively studied as it defines the key challenge towards large scale industrial application [36, 37]. This dissertation therefore aims for a deeper understanding of the most important reliability aspects and for possible ways to overcome these obstacles. Under the term reliability, three major aspects can be defined, being

- **Variability**,
- **Retention**, and
- **Endurance**

which are therefore covered in the three main chapters of this work. **Variability** means the stochastic nature of both the switching process as well as the short term instability of programmed states. Whereas variations between consecutive switching cycles can be overcome by appropriate programming algorithms, random fluctuations of the programmed states are identified as one of the key challenges for the application of ReRAM devices, as will be laid down in chapter 4.
Besides this short-term instability, the long term non-volatility or **retention** is one of the key features of ReRAM devices. Chapter 5 will tackle the long-term stability of the programmed states. Here, different model approaches will be presented which explain the observed degradation effects. Since retention targets like 10 years at operating temperatures of (e.g.) 85 °C (given by the respective application) cannot be measured directly, a model approach is developed which allows for extrapolation of data measured at higher temperatures towards the target retention time at lower operating temperature. One of the key findings will be the interconnection of variability and retention.
Finally, the number of possible switching cycles until device failure, called **endurance** is investigated in chapter 6. Here, a sophisticated programming algorithm will be presented which allows for fast endurance assessment in combination with adaptive programming conditions to maximize the device endurance. Furthermore, appropriate material selections in order to maximize the endurance are discussed.
In industrial application millions and higher orders of magnitudes of devices will be used. Even if the majority of devices functions within acceptable reliability margins, the few worst cells will be the limiting factor for the whole memory array. Therefore the investigation of single or typical cells (a single cell has a high chance to behave like the median) is inappropriate to determine and improve reliability aspects. Instead, it is crucial to find the reliability of the full statistics in order to predict the behavior of all cells. By analyzing the internal statistics of the observed quantity, reliability aspects may be characterized by changes in the statistical parameters. Thus, this dissertation focuses strongly on the characterization of large sets of ReRAM devices. Firstly, this enables statistically sound results for the three tackled reliability aspects variability, retention and endurance. Secondly, and more importantly, this approach allows to identify challenges like changes in intrinsic statistics which would be undetectable in single or few cells.

2 Fundamentals

This chapter provides a brief overview of the fundamental aspects building the foundation of the findings in this work. Firstly, an introduction to redox based resistive switching memories is given. In particular, resistive switching memories based on the valence change mechanism are introduced. Secondly, the state of the art of the reliability aspects investigated in this work is outlined.

2.1 Classification of ReRAM

In 1971, Leon Chua proposed that besides resistor, capacitor and inductor, a fourth fundamental electrical component should exist which links electric charge and magnetic flux [38]. Later, the theory was generalized and it was found that resistive switching random access memory (RRAM) fills the gap discovered by Chua [39, 40]. In general, RRAM is a class of non-volatile memory which utilizes changes of electrical resistivity to store information [10]. Eminent types of RRAM are magnetoresistive random access memory (MRAM), phase change memory (PCM) and redox based resistive switching memory (ReRAM) [11]. Whereas MRAM relies on the orientation of the polarization in a magnetic tunnel junction (MTJ) [41], PCM utilizes the transition between amorphous and crystalline phases [42]. This dissertation focuses on ReRAM which resorts to internal redox reactions to reversibly switch the resistivity of the active material [13].
Different types of ReRAM devices can be distinguished concerning the operating voltage polarities as sketched in figure 2.1. In general, ReRAM devices can be switched between at least two different resistance states. The process of switching the devices into the low resistive state (LRS) is called SET and the reversed operation resulting in the high resistive state is called RESET. In unipolar switching devices, as depicted in figure 2.1 a), both the SET and RESET process can take place in a single voltage polarity. Here, usually the SET voltage is higher than the RESET voltage [11]. However, the cell current during the SET process needs to be limited to prevent a thermal run-away as soon as the resistance decreases. Thus, a current compliance (CC) is applied during SET, limiting the current to I_{CC} which is typically lower than the peak current during RESET [43, 44]. A prominent example for unipolar switching

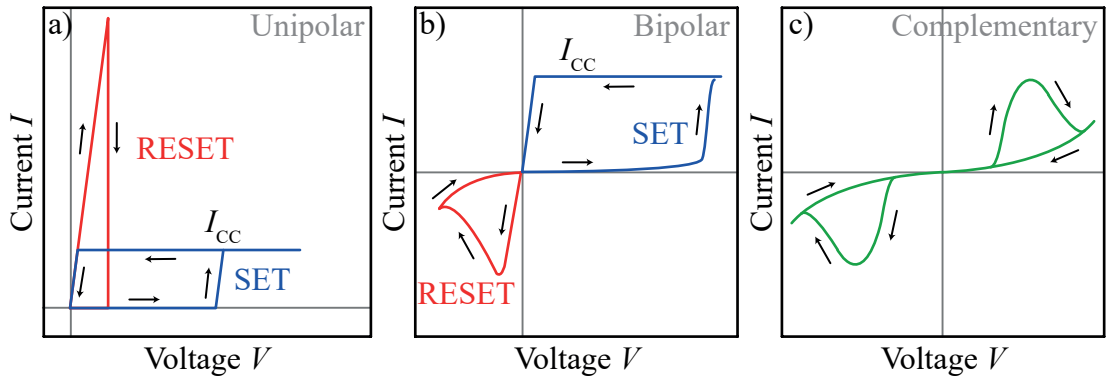

FIGURE 2.1: $I - V$ schematics of different ReRAM switching modes. a) Unipolar switching with SET and RESET being independent of the voltage polarity, b) Bipolar switching with opposing SET and RESET polarity and c) Complementary switching. Redrawn from [11].

devices are thermochemical memories (TCM). These devices typically consist of a symmetrical metal / metal-oxide / metal stack. A resistance change is realized by applying appropriate voltages which induce local temperature gradients leading to redox reactions and variations in the local stoichiometry [10]. The TCM mechanism is explained in more detail in [45–47].
In bipolar switching devices, as shown in figure 2.1 b), SET and RESET voltages need to be applied in opposing polarities [48]. Here, the RESET current can be lower than the current compliance in the SET operation. In contrast to unipolar devices, no current compliance is required in some bipolar systems [10]. The mostly recognized examples for bipolar switching ReRAM are

- electrochemical metallization memory (ECM), and
- valence change memory (VCM) [10, 49].

ECM, also called conductive bridge RAM (CBRAM) or programmable metallization cells (PMC) is based on the electrochemical dissolution of an active electrode metal. The dissolved metal ions migrate towards in inert counter electrode through an ion conducting electrolyte layer sandwiched between both electrodes [10]. In the initial state the conductivity of the cell is limited by the low electronic conductivity of the electrolyte. During the SET operation, the generation and migration of metal ions results in the formation of a conductive bridging filament resulting in a low resistive state. Dissolving this filament by applying a RESET voltage in the opposite polarity switches the cell back to the high resistive state [10, 50, 51]. The ECM mechanism is laid down in more detail in [52–55]
VCM, in literature also recognized as oxide based RAM (OxRAM), is based on a partial reduction of a switching transition metal oxide and will be explained in detail

in the next section (see also [10]).
Besides unipolar and bipolar switching, a complementary resistive switching (CS or CRS) mode is reported [56–59] and displayed in figure 2.1 c). CRS cells can either be obtained by anti-serial combination of two bipolar switching cells [56] or by fabrication and operation of single cells with suitable parameters [58]. The cells appear to be always high resistive when measured at low read voltages since both switching operations end up in the high resistive state. By reading at higher voltages a current increase is observed if the cell was programmed with the opposite polarity before. Otherwise, the cell appears to be high resistive. In contrast to unipolar and bipolar switching, the read operation in CRS cells is typically destructive, i.e. the state of the cell has to be reprogrammed after reading [60]. However, CRS switching devices have been proposed as a possible solution for sneak path or selector problems in passive ReRAM arrays [56, 58].

2.1.1 VCM

VCM type resistive switching devices typically consist of a MIM layer structure. Here, the insulating layer I typically is a thin film (< 10 nm) of a binary transition metal oxide like HfO_x, ZrO_x, TiO_x or TaO_x [13, 14]. Additionally, ternary oxides like $SrTiO_x$ are used [10]. The metal electrodes M are usually comprised by one active electrode (AE) and one ohmic electrode (OE). For the active electrode a metal with high work function is used (e.g. Pt, Ir, TiN) which forms a Schottky barrier at the interface between metal and oxide. The ohmic electrode typically consists of a metal with lower work function and higher oxygen affinity, forming an ohmic contact at the interface [10].
The general working principle of VCM type ReRAM is a partial reduction of the cation sub-lattice by partial oxidation of the OE. For a simpler description, the remaining “holes” created by each excorporated oxygen ion are called oxygen vacancies. According to the Kröger-Vink denotation [61], one oxygen vacancy has the symbol $V_O^{\cdot\cdot}$, reading vacancy (V) at an oxygen (O) lattice site. Here, the two dots identify the oxygen vacancy as two-fold positively charged with respect to the stoichiometric oxide. However, electrically the oxygen vacancy remains neutral because the local charge will immediately be neutralized by mobile electrons. Due to these mobile electrons, each $V_O^{\cdot\cdot}$ can be considered a donator dopant which is utilized to decrease the resistance of the switching oxide.

Electroforming

Prior to resistive switching, the initially very large resistivity of pristine VCM cells needs to be lowered by an electroforming (or forming) operation [10]. A comparatively high negative voltage is applied to the active electrode which leads to the generation of oxygen vacancies at the interface of oxide and ohmic electrode. According to the Kröger-Vink denotation [61], the generation of one oxygen vacancy $V_O^{\cdot\cdot}$ can be written as

$$O_O^X \rightarrow V_O^{\cdot\cdot} + 2e' + \frac{1}{2}O_2, \tag{2.1}$$

where O_O^X is an oxygen ion on an oxygen site [62]. The free oxygen atom ($1/2O_2$) can either be evaporated as oxygen gas [63], or participate in partial oxidation of the ohmic electrode metal. The generated $V_O^{\cdot\cdot}$ lead to a decrease of the oxide resistance and therefore increased current [64]. The latter causes a temperature increase in the oxide due to Joule heating which accelerates the generation of oxygen vacancies. This leads to a self-reinforcing thermal run-away which may destroy the cell if the current was not limited by a current compliance, series resistor or transistor [62]. Historically, this dielectric breakdown was considered a major reliability issue in metal oxide field effect transistors (MOSFET) since it could affect the gate dielectric [65]. However, a controlled dielectric breakdown leads to the formation of an *n*-conducting filament, lowering the resistance of the switching oxide [62]. Though, it may be noted that besides filamentary VCM, also interface-type switching is reported. Instead of accumulating oxygen vacancies in a confined filament, they distribute over the whole cell area [10]. This work will mainly focus on the filamentary type VCM.
In general, electroforming can be performed by either a triangular voltage (or current) sweep, or by applying a rectangular pulse. In both cases, a certain threshold has to be reached. Using a triangular sweep, the electroforming will occur at a certain threshold voltage (depending on the sweep rate). In the pulsed operation the probability to initiate electroforming increases with the pulse length and height. These forming kinetics depend on several conditions [62]. Besides the internal micro-structure (e.g. grain size) and material properties of the switching oxide [62], exterior influences can have a high impact on the forming kinetics. As expected intuitively, the ambient temperature determines the forming kinetics to a certain degree [62, 66]. Furthermore, it was demonstrated that the ambient atmosphere plays a major role. On the one hand, the oxygen partial pressure seems to impact the resistance of the switching oxide [62, 67]. On the other hand atmospheric moisture has been reported to have a huge impact on the forming kinetics [63].
Additionally, the material of the OE has a direct impact on the forming kinetics. Guo et al. calculated the defect formation energy which is required to generate one oxygen

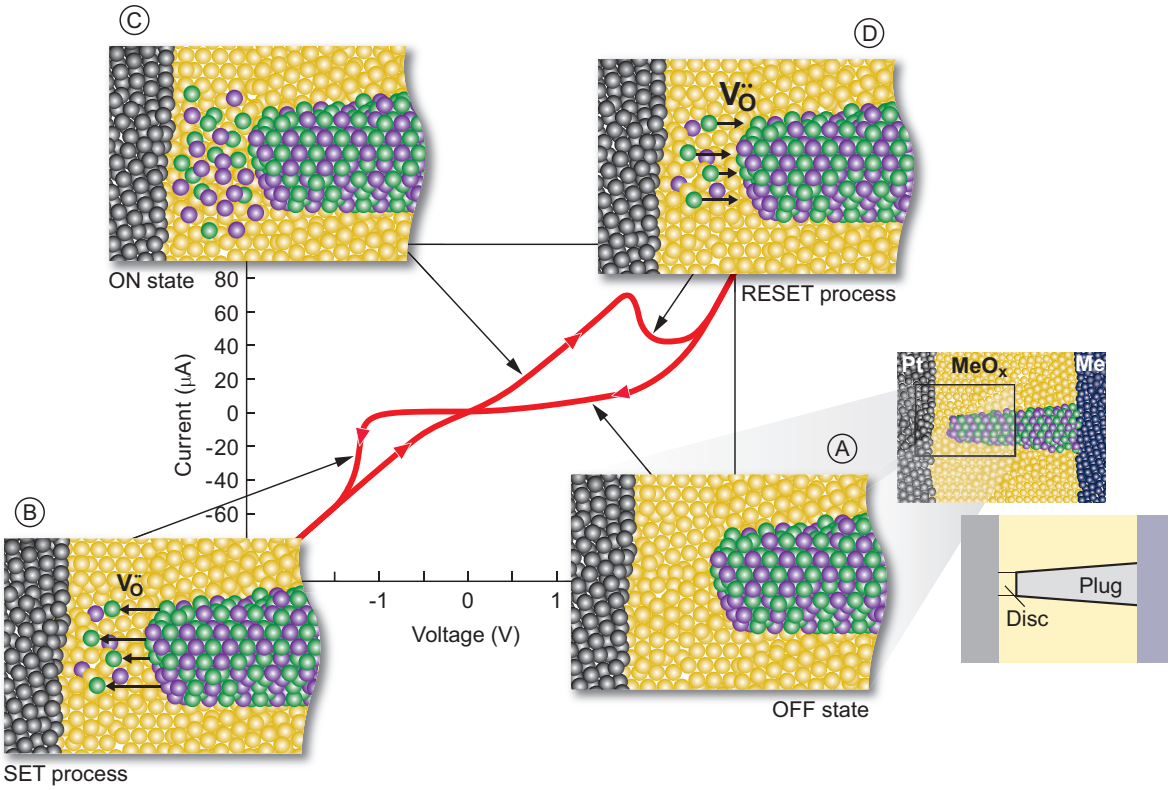

FIGURE 2.2: Schematic of the (counter eight-wise) VCM switching mechanism for an exemplary Pt / ZrO_x / Zr cell. The conductive filament consists of *plug* and *disc* (inset on the right). (A) HRS or OFF state. The oxygen vacancies in the *disc* are depleted, resulting in a high Schottky barrier and thus high resistance. During the SET process (B) oxygen vacancies are shifted from *plug* to *disc* due to a negative voltage at the AE, resulting in the LRS or ON state (C). The low resistance is explained by a lowered Schottky barrier induced by the increased concentration of oxygen vacancies in the *disc*. A positive voltage at the AE shifts oxygen vacancies back into the *plug* and RESETs (D) the cell. Reproduced with permission from [10], © 2012 Wiley-VCH Verlag.

vacancy at the interface between OE and oxide [7]. Therefore, they determined the formation energy of one oxygen vacancy in the bulk oxide and subtracted the oxide free energy gained by oxidizing the respective OE metal [7]. As a result, it is reported that the defect formation energy increases with the oxygen chemical potential of the metal and thus the forming voltage increases [7].

Resistive Switching

Figure 2.2 illustrates the general switching characteristics of formed VCM cells at the example of a Pt / ZrO_x / Zr cell. The active Pt electrode is shown on the left and the ohmic Zr electrode frames the switching ZrO_x from the right.
After electroforming, the cell is typically in the LRS or ON state, shown in figure 2.2, (C). The metal oxide layer is bridged by an *n*-conducting filament induced by oxygen vacancies. This filament consists of a *plug* region attached to the ohmic electrode and

a *disc* region in the vicinity of the active electrode. Here, the Schottky barrier height at the AE/oxide interface is strongly influenced by the oxygen vacancy concentration of the *disc*. In the LRS (ON) this concentration is high resulting in a low Schottky barrier and thus a low resistance of the cell [10].
The application of a positive voltage to the active electrode leads to a retraction of the oxygen vacancies from the *disc* into the *plug*. This RESET process (D), locally re-oxidizes the disc and thus increases Schottky barrier and resistance. As a result, the cell is in the HRS or OFF state shown in figure 2.2, (A). It may be noted that the resistance in the HRS is usually significantly lower than the initial resistance of the cell in its pristine state [10]. The SET process (B) is initiated by application of a negative voltage to the active electrode which shifts oxygen vacancies from the *plug* to the *disc* and restores the LRS or ON state (C) [10].
The schematic shows the most common switching mode of VCM cells. Apart from this so called counter eight-wise switching (c8w, because of the $I - V$ characteristic following the reversed drawing direction of the number eight [10]), also eight-wise (8w) or anomalous switching is reported [68]. Although both switching modes can occur in several VCM systems [69–72], this dissertation focuses on the more common c8w switching. Despite the fact that the presented VCM switching model is suited to explain several experimentally observed VCM characteristics [73], the conduction mechanism in VCM-type ReRAM is still under debate. Alternative approaches to model the VCM mechanism are the hour-glass model [74] or the quantum point contact (QPC) model [75–77]. With respect to the inherent variability of several VCM systems, especially models with a focus on the trap-assisted tunneling mechanism come into account [5, 78, 79].

2.2 Reliability Aspects

The preceding section introduced VCM type switching ReRAM and covered the general characteristics of these systems. With the step towards industrial application of VCM cells, several challenges occur [37, 80]. The scope of this dissertation is to gain a better understanding of the reliability aspects of VCM devices. Therefore, the state of the art of the most relevant reliability aspects is laid down in the following sections.

2.2.1 Variability

One of the most critical aspects regarding the reliability of VCM cells is its stochastic nature [37, 81]. To characterize this stochastic nature, first of all different types of

variability can be distinguished:

- Device to device (D2D) variability,
- Cycle to cycle (C2C) variability, and
- Read to Read (R2R) variability [82].

D2D variability covers stochastic variations between single VCM cells. Here, it may be further distinguished between variations of cells on the same die, between multiple dies of the same wafer or even from wafer to wafer. Based on D2D variability, the forming voltage or time varies between single cells [83–85]. However, once forming was successful, it has been demonstrated that stochastic variations from one switching cycle to the next (C2C) are usually larger than the D2D variability [86, 87].

Cycle to Cycle Variability

As explained above, the switching process of VCM ReRAM devices is based on the migration of oxygen vacancies. This ionic defect migration is reported to exhibit a rather stochastic nature [88, 89]. On the one hand, this results in a variability of the programmed states between switching cycles [90]. On the other hand, the switching kinetics (i.e. SET/RESET voltage and time) varies from cycle to cycle [91, 92].
Figure 2.3 illustrates C2C variability for ZrO_2 based VCM ReRAM devices. Figure 2.3, a) shows exemplary *I-V*-sweeps measured on ZrO_2 based VCM ReRAM. It is observed that the HRS resistance evaluated prior to the SET event significantly varies between subsequent sweeps. Analogously, the LRS exhibits fluctuations between switching cycles. However, the variability of the HRS is significantly higher compared to the LRS. This observation is confirmed by the exemplary endurance measurement shown in figure 2.3, b). While cycling up to 7 M cycles, the state of the cell is read by a 0.2 V pulse every 500 SET and RESET cycles. The resulting read current is depicted by the circles in figure 2.3, b) and observed to vary within a certain band. Again, the HRS resistance varies in a broader range than the LRS.
Thus, within C2C variability it can be distinguished between SET [94, 95] and RESET [96–98] variability [90]. Despite the usually higher variability of the RESET operation [85], it has been demonstrated that SET and RESET variability are not independent and may affect each other [90, 95, 99]. Additionally, the variability of the switching kinetics is affected by the C2C variability of the resistance. This is also observed in figure 2.3, a). With increasing HRS read current (decreasing resistance) the SET voltage decreases.
To ensure a sufficiently large read window over the full life span of a VCM cell (endurance), a reduction of the C2C variability is highly desirable. Besides tuning

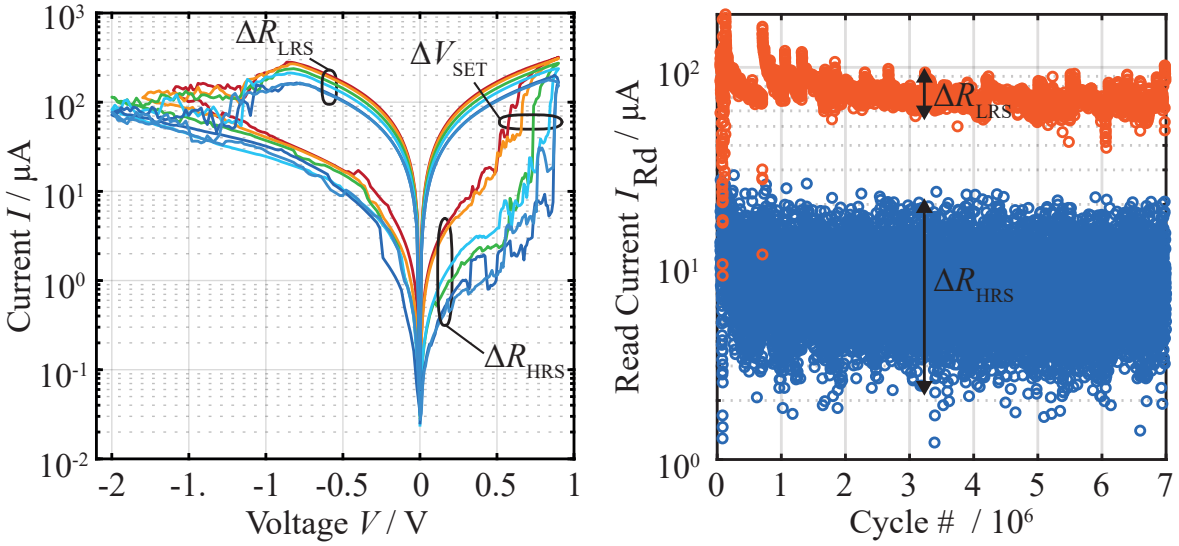

FIGURE 2.3: C2C variability illustrated by measurements on $Pt/ZrO_2/Ta/Pt$ devices. a) Exemplary *I-V*-sweeps. Both HRS and LRS resistance vary between subsequent cycles. With HRS resistance also the SET voltage varies. b) Endurance measurement with SET voltage V_{SET} = 0.9 V, RESET voltage V_{RESET} = -1.4 V, read voltage V_{RD} = 0.2 V (all applied to Ta electrode, Pt electrode connected to GND) and pulse length $V_{SET/RESET}$ = 20 µs. HRS and LRS vary in a certain band which is broader for HRS. Data taken from [93].

of the device structure and material stack [100], one potential approach to influence the C2C variability is the forming operation [101]. It was demonstrated that a higher forming current may result in a larger conductive filament radius which is suggested to be beneficial for a reduction of the variability [101]. Additionally, program verify algorithms can be used to reduce the impact of C2C variability and to widen the read window [89]. Here, the cell resistance is evaluated after each programming operation. Thus, additional programming pulses can be applied if the cell exceeds a defined threshold.

Read to Read Variability

Although such programming schemes counter act C2C variability, they are reported to have a limited impact since the accompanying shaping of the read current (or resistance) distribution has no lasting effect [86, 102]. This is explained by fluctuations of the programmed states which do not necessarily occur during switching operations but also without applied bias. Thus, variability is also observed from read to read (R2R) which limits the impact of program verify (or shaping) algorithms [86, 102, 103].

One frequently used term in the context of R2R variability is random telegraph noise (RTN) which is characterized by random jumps of the observed quantity (e.g.

read current) between rather discrete levels [104–106]. The origin of these jumps is often attributed to the trapping and de-trapping of electrons at defect states in the oxide [107–110]. Alternative explanations are based on the random reconfiguration of defects (e.g. oxygen vacancies) in the measured oxide [105, 111], Besides the observation of RTN in ReRAM, it has been extensively studied as reliability issue of the gate dielectric in MOSFETs [112–114]. However, the nomenclature regarding R2R variability and RTN is not entirely explicit. In many cases different modes of R2R variability are observed in the same programmed state. Often a high frequent fluctuation with low amplitude between two states is accompanied by lower frequent jumps with significantly higher amplitude. In some publications both processes are covered by RTN [105, 109]. Other publications distinguish between RTN and ionic noise [111], random walk [115] or base line jumps [103].
In general, it is observed that R2R variability strongly depends on the resistance of the investigated state [82]. Typically, R2R variability is less pronounced in the LRS compared to the HRS [116], given that the LRS is sufficiently low resistive. In case of a higher resistive LRS, it may be noted that HRS and LRS variability can become equal [117]. Within different HRS states, a transition from low noise at high resistances (MΩ) towards high noise at lower resistance (kΩ) is reported [96]. Besides the resistance, the preceding programming operation may affect the observed R2R variability [118].
Whereas the changes in the noise characteristics at different resistances and under different programming conditions may provide a deeper understanding of the underlying conduction mechanisms [119, 120], R2R variability is one of the key challenges in the reliability of ReRAM devices [86]. With increasing number of cells in a memory device, R2R fluctuations result in a broadening of the resistance distribution which limits the resulting read window [121–124].
For the evaluation of experimentally obtained noise characteristics, two concepts are commonly used. On the one hand, the power-spectral-density (PSD) of the signal can be calculated and plotted versus the frequency f of the fluctuations [125, 126]. The resulting plot usually shows a slope following $1/f^a$. The value of the exponent a which best fits the data gives rise to the underlying type of noise. Typically observed are

- white noise with $a = 0$,
- pink noise with $a = 1$, or
- brown noise with $a = 2$ [127].

Another approach to analyze R2R variability or RTN is the use of factorial hidden Markov models (FHMM) [128] which will be covered in section 2.3.

TABLE 2.1: Summary of reported retention characteristics. Adapted from [82] and extended

Source	Material	LRS current	HRS current
IMEC [36, 129–131]	HfO_2/Hf	Decreasing	Increasing/Decreasing
Panasonic [132, 133, 142]	TaO_x	Decreasing	Increasing
Macronix [103]	HfO_2 $HfAlO_x$	Stable	Increasing
Laboratorio MDM [134]	WO_x	N/A	Increasing
Politecnico di Milano [135]	NiO	Decreasing	N/A
Stanford University [136]	HfO_2	Decreasing	N/A
IHP Frankfurt [137]	HfO_2	Decreasing	N/A
LETI [138, 139]	$Hf(Al)O_x$ TaO_x	Decreasing	N/A
Fudan University [140]	CuSiO	Decreasing	N/A

2.2.2 Retention

The retention of a memory device describes the long-term stability of a programmed state. Typical requirements for NVM applications of ReRAM are 5-10 years at 85-125 °C [82]. Within this time at the respective operating (or storage) temperature, the HRS or LRS has to be stable enough to ensure a sufficiently large read window. Several groups investigated the retention of VCM ReRAM [30, 103, 129–141]. As summarized in table 2.1, they typically find that the LRS read current decreases over time. Usually, this effect is attributed to a loss of oxygen vacancies and degradation of the conductive filament resulting in a higher resistance of the cell [36, 129–131]. The retention of the HRS is covered less in literature. Here, it is reported that the read current either decreases or increases over time [36, 103, 130–133]. A current decrease in the HRS state is explained analogously to the LRS by the loss of oxygen vacancies [130]. A current increase in the HRS is explained as diffusion of oxygen vacancies from the filament towards the depleted region (gap or disc) close to the active electrode [132].

Whereas several studies focus on the development of typical cells or the median of the distribution [129, 139, 143], other studies demonstrate that the distribution tails (or weak bits) may show a completely different behavior [36]. However, the most promising approaches to explain retention characteristics are based on statistics and cover the development of the whole distribution [144].

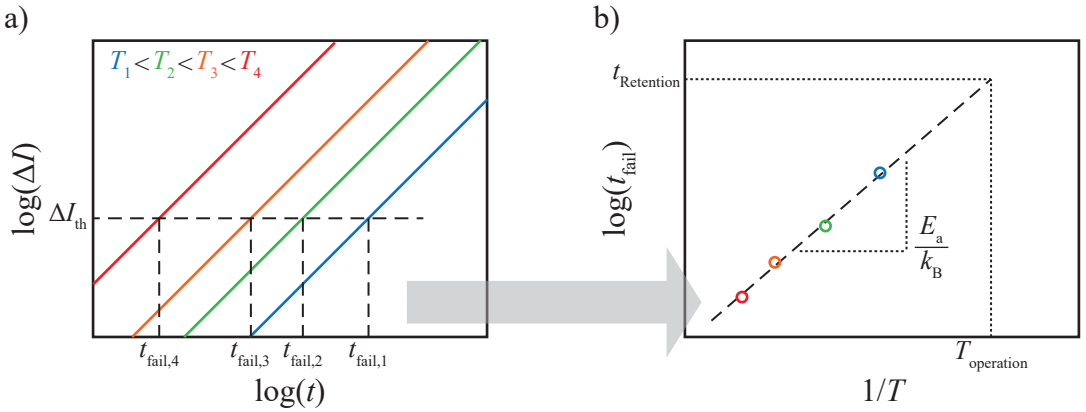

FIGURE 2.4: a) Schematic degradation of a programmed state over time for different baking temperatures (colors). With increasing temperature, the failure criterion ΔI_{th} is reached earlier. Ideally, the slope of the degrading current is parallel on the double logarithmic scale, i.e. constant activation energy regardless of the definition of ΔI_{th}. b) The times $t_{th,i}$ to reach the failure criterion ΔI_{th} are transferred to an Arrhenius plot to extract the activation energy E_a and to extrapolate the retention time $t_{Retention}$ for the required operating temperature $T_{operation}$.

Activation Energy

Whereas the endurance (c.f. section 2.2.3) of ReRAM devices can be easily tested by fast cycling up to the required benchmark, retention requirements of 10 years at operating temperature cannot be verified directly. Instead, retention experiments are typically conducted at elevated temperatures (also called accelerated life testing, ALT). Appropriate models have to be developed to extrapolate the obtained data to longer storage times at lower temperatures. The most common approach is to determine an activation energy from the electrical data for the underlying degradation processes [130, 135, 137–139, 145–147]. Typically, a failure criterion for the read current or resistance is defined. This may be exceeding a certain threshold current [137] or a resistance change by a defined factor with respect to the initial resistance [138, 139]. The time t_{fail} to reach the defined failure criterion (often *mean time to failure* MTTF) is determined for different temperatures, as schematically shown in figure 2.4, a). If the observed degradation is based on a temperature activated effect, the time to failure t_{fail} should follow an Arrhenius equation, reading

$$t_{fail} = A \cdot \exp\left(\frac{-E_a}{k_B T}\right), \tag{2.2}$$

with a pre-exponential factor A, the activation energy E_a, the Boltzmann-constant k_B and the temperature T [148]. By plotting $\log(t_{fail})$ versus $1/T$, as shown in figure 2.4, b), E_a can directly be determined from the slope of the so called Arrhenius plot. At the same time, this slope can be extrapolated towards the desired retention time $t_{Retention}$

(e.g. 10 years) at the required operating temperature $T_{\text{operation}}$ (e.g. 85-125 °C).
However, this approach has some disadvantages: First of all, it is not clear if a change of the observed quantity (ΔI) is directly correlated to an underlying, temperature activated process. Multiple processes with varying activation energies may occur with the superposition of all processes resulting in the observed electrical change. Even if only diffusion of oxygen vacancies is taken into account, first principles simulations suggest varying activation energies within the same material [121, 149]. In this case, the activation energy can only be extracted for the rate limiting process which may change in the transition from elevated test temperatures towards lower operating temperatures.
Furthermore, the choice of the failure criterion may have an impact on the determined activation energy. Ideally, the slopes of the degradation are parallel for different temperatures on the double-logarithmic scale, as demonstrated in figure 2.4, a). In this case, E_{a} is independent of the failure criterion. However, the reported degradation usually deviates from the ideal trend [82, 129, 134, 141] and thus the obtained activation energy depends on the defined failure criterion. Additionally, the percentage of cells to reach the criterion affects the activation energy. The different degradation characteristics of median and distribution tails, mentioned before, will result in different values of E_{a}.
In conclusion, the Arrhenius approach is a simple and widely accepted method to extrapolate high temperature retention experiments towards operating temperatures. Nevertheless, more sophisticated models are required for a reliable estimation of the long-term stability of the programmed states. One alternative approach will be demonstrated in chapter 5.2.

2.2.3 Endurance

In the context of memory devices, endurance is the maximum number of switching cycles (1/0, SET/RESET) a device can perform until failure. With the typical endurance of $10^4 - 10^5$ cycles for flash and up to 10^{15} cycles for DRAM, a huge endurance gap emerged which may potentially be filled by ReRAM [82]. Over the last years, several groups developed VCM ReRAM devices with endurance of $10^7 - 10^{12}$ cycles as summarized in table 2.2 [16, 26, 129, 139, 150–159]. It may be noted that the majority of collected publications only demonstrate endurance on single or few devices. In contrast to retention, only few publications provide a statistical endurance evaluation of a higher number of cells [152, 158].
A typical endurance plot is shown in figure 2.3, b). A Pt / ZrO_2 / Ta / Pt cell is cycled by 20 µs pulses with a SET voltage of 0.9 V and a RESET voltage of -1.4 V, applied to

TABLE 2.2: Summary of reported endurance characteristics. Adapted from [82] and extended

Source	Material	SET	RESET	Cycles
SAIT [16, 26]	TaO based	4.5 V, 10 ns	7 V, 10 ns	10^{12}
ITRI [150, 151]	HfO_2/Ti	3.2 V, 40 ns	2.7 V, 40 ns	10^{10}
LETI [152]	HfO_2/Ti	2 V, 500 ns	2.5 V, 500 ns	10^{7}
LETI [139]	HfO_2/Ti	2 V, 0.5 µs	1.4 V, 10 µs	10^{8}
IMEC [129, 153]	HfO_2/Hf, Ti	1.8 V, 5 ns	1.8 V, 10 ns	10^{10}
HP [154]	TaO based	1.9 V, 1 µs	2.2 V, 1 µs	$1.5 \cdot 10^{10}$
Panasonic [155, 156]	TaO based	1.5 V, 100 ns	2 V, 100 ns	10^{9}
SEMATECH [157]	HfO_x	1.5 V, 50 ns	1.5 V, 10 ns	$> 10^{9}$
Wuhan [158]	AlO_x	1.05 V, 28 ns	1.25 V, 28 ns	10^{10}
Hsinchu [159]	WO_x/ZrO_x	3.7 V, 100 ns	5 V, 100 ns	10^{8}

the Ta/Pt electrode. The read current of HRS and LRS at 0.2 V is evaluated every 500 cycles and plotted as circle. As discussed before, significant R2R fluctuations occur which have to be taken into account in endurance evaluations. However, the overall trend of HRS and LRS is comparatively stable within the tested 10^7 cycles.
Here, the experiment was terminated before the endurance failure occurred. Depending on material and programming conditions, this failure may be caused by either drifting of the HRS towards the LRS or vice versa. Eventually, the device becomes stuck in one of the states. Thus, it was demonstrated to be crucial to balance SET and RESET pulse width and height to maximize the endurance [36, 153]. A too strong RESET operation would cause the system to lose oxygen vacancies by recombination with oxygen from the electrodes until the SET fails. Vice versa, an excessive SET pulse may cause the generation of too many oxygen vacancies until the RESET fails [82, 153]. Besides balancing fixed programming conditions, algorithms with adaptive SET and RESET pulses have been reported to improve the endurance [160].
Even with balanced programming conditions, endurance failure may occur. One reason could be local material changes in the filament or surrounding oxide [82]. The thermal stress during cycling may induce local modulations like crystallization which may increase the resistance or SET voltage [131, 161]. It was furthermore demonstrated that a low power consumption of the switching operations may prolong the endurance of a system [162, 163]. Thus, Nail et al. proposed a maximum energy E_{max} the dielectric can sustain until failure [164]. E_{max} is proportional to the

number of moved oxygen vacancies per cycle $N_{\text{moved per cycle}}$, the energy barrier to move one oxygen vacancy E_{barrier} and the number of cycles N_{cycles}, reading

$$E_{\text{max}} \propto N_{\text{moved per cycle}} \cdot E_{\text{barrier}} \cdot N_{\text{cycles}} \quad [164]. \tag{2.3}$$

With $\text{N}_{\text{moved per cycle}}$ being related to the read window, E_{barrier} determining retention and N_{cycles} this leads to a trade-off of read-window, retention and endurance [164] which is reported by several groups [129, 139, 141].

Materials Selection

Apart from the optimization of the programming conditions, the device stack and materials can be tuned towards higher endurance [165–167]. Since RESET failures are often induced by generation of excessive oxygen vacancies during cycling, Guo et al. proposed to tune the energy barrier for oxygen vacancy generation [7]. They calculated the electronic structures of typical oxides for VCM ReRAM via the ab-initio plane wave pseudopotential method using CASTEP code. By this, the defect formation energy for the generation of one oxygen vacancy in the bulk oxide $E_{\text{form,bulk}}$ is determined [7]. This barrier is comparatively high and the generation rather unlikely. However, in the vicinity of an oxygen scavenging layer (ohmic electrode) the barrier is reduced by the oxygen chemical potential of the respective electrode metal μ_{Me}. Thus, the defect formation energy E_{form} for the generation of one oxygen vacancy at the oxide/electrode interface reads

$$E_{\text{form}} = E_{\text{form,bulk}} + \mu_{\text{Me}} \quad [7]. \tag{2.4}$$

For example, the energy for the formation of an oxygen vacancy in bulk HfO_2 of 5.9 eV is reduced to 0.1 eV with Hf ($\mu_{\text{Hf}} = -5.8$ eV) as oxygen scavenging layer. In contrast, a Ru electrode with $\mu_{\text{Ru}} = -1.8$ eV would result in a defect formation energy of $E_{\text{form}} = 4.1$ eV. Following Guo et al. the higher formation energy should lower the probability of RESET failures [7]. The general trend was already demonstrated experimentally for Ta_2O_5 based devices by Kim et al. [168]. A statistics based study of the impact of the defect formation energy on endurance and forming voltage of ZrO_2 based devices is given in chapter 6.

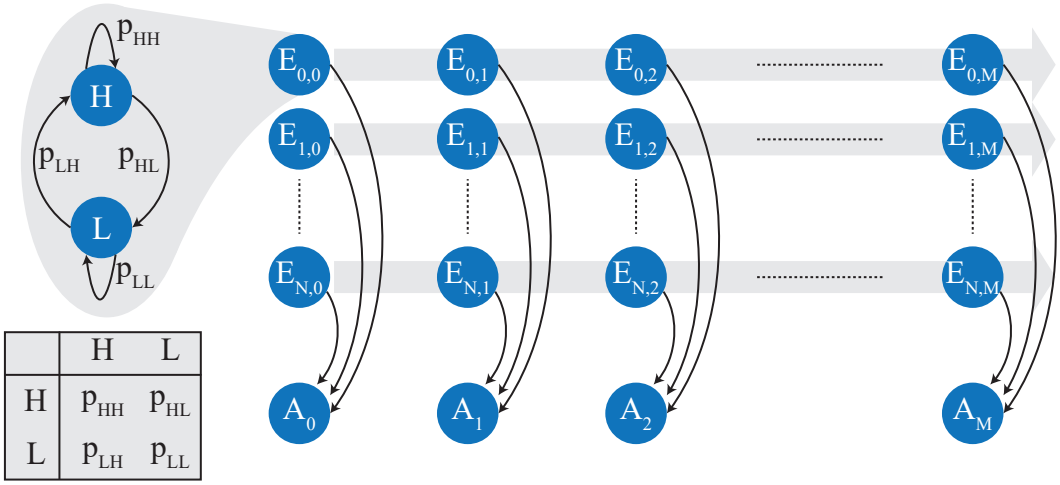

FIGURE 2.5: Schematic of an FHMM. One Markov chain (shown on the left) is characterized by two hidden states (L and H) and the probabilities p_{HH}, p_{HL}, p_{LH} and p_{LL} for the transitions between L and H. The FHMM contains N individual Markov chains which generate an output A_m at each instant of time m. The outputs $A_{0...M}$ can be observed as e.g. noisy current signal. The underlying (hidden) Markov chains E need to be estimated by appropriate fitting algorithms. Adapted from [128].

2.3 Factorial Hidden Markov Models

Hidden Markov models [169] and factorial hidden Markov models [170] were proposed by Puglisi et al. as a powerful tool for the evaluation of noise (RTN) characteristics in ReRAM devices [108, 128, 171]. Both HMM and FHMM are based on Markov chains. The difference between the two approaches is the utilization of single chains in HMM and the concatenation of multiple chains in FHMM. Each Markov chain is an element which can fluctuate between a defined number of discrete levels (hidden states). The element is defined by its hidden states and a matrix containing the transition probabilities between those states. A schematic Markov chain is shown on the left side of figure 2.5. The drawn Markov chain has the two hidden states L and H which could be considered one low (L) and one high (H) current state of a noisy read current trace. Besides these current levels, the Markov chain is characterized by the probabilities p_{HH}, p_{HL}, p_{LH} and p_{LL} for the four possible transitions indicated by black arrows.

In HMM, the depicted Markov chain could be used to model a 2-state RTN signal which fluctuates only between two discrete levels [128]. To model multi-level RTN, a Markov chain with multiple hidden states could be considered. However, FHMM was demonstrated to be a superior approach for multi-level RTN signals [128]. Here, multiple (usually 2-state) Markov chains are considered and the superposition of all Markov chains results in the modeled noise signal, as sketched in figure 2.5. At each instant of time (0-M), the output A (observable current) is generated by superposition

of the hidden states of all considered Markov chains E. Fitting an experimental noise signal with an appropriate FHMM allows to decompose the signal into single (2-state) components [171]. Subsequently, the individual components may be attributed to single traps being the origin of the observed R2R variability.
In general, it is easy to generate artificial RTN by defining an FHMM consisting of N Markov chains E_{N} and calculating the most likely output A_{m} at each increment of time m. However, the more interesting operation is to find the appropriate FHMM to describe a measured signal. Respective algorithms are used to determine the most likely set of Markov chains for a given noise signal. Prominent examples are the Viterbi-algorithm [128, 169, 172] and the Baum-Welch algorithm [173, 174].
Although this approach is accompanied by a comparatively high computational burden, it offers several advantages: It provides comprehensive information about the single fluctuating components and intrinsically returns the number of components [171]. Furthermore, the algorithms can be well parallelized, since the convergence to the right solution strongly depends on initial parameters. Thus, multiple instances of the algorithm with different initial values may be performed on different computer cores.

2.4 Kinetic Monte Carlo Model

To understand the stochastic nature of the current transport and switching characteristics in VCM ReRAM, several kinetic Monte Carlo Models have been published [175–179]. The statistical models presented in this work (c.f. chapter 4.3.1 and 5.3) are based on the kinetic Monte Carlo model developed by Abbaspour et al. [3–6, 180]. Thus, the general concept and the most important equations of the model will be laid down in this section.
The model is based on a cubic box of the dimensions 5 nm x 5 nm x 5 nm which represents an HfO_2 layer. The box is discretized by a mesh of 21 x 21 x 21 lattice sites, resulting in a resolution of 0.25 nm. As sketched in figure 2.6, a), each lattice site can either be considered as oxide matrix or contain an oxygen vacancy. Here, the oxygen vacancies induce defect states in the band gap of the partially reduced oxide which serve as electron traps and support the electrical transport in the material [6]. Apart from these traps, no other defects are implemented in this model [6]. In one dimension (z) the oxide is sandwiched between the active (AE) and the ohmic (OE) electrode. Throughout this work, the voltage is applied to the ohmic electrode and the active electrode is connected to ground.
The flow chart in figure 2.6, b) provides an overview of the relevant simulation steps. In each increment of time, the potential at each lattice site is calculated, followed

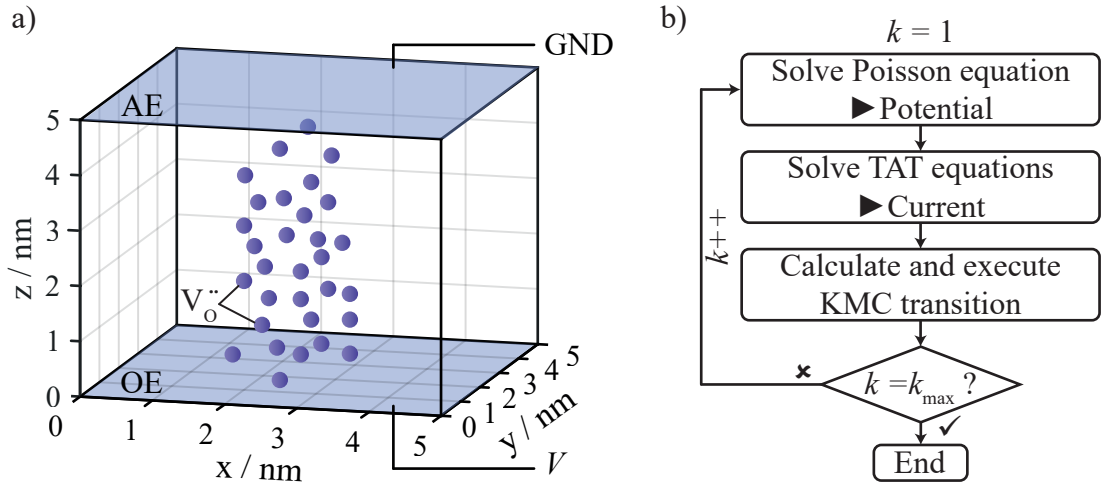

FIGURE 2.6: a) Structure of the KMC model. A cubic oxide mesh is sandwiched between active (AE) and ohmic (OE) electrode. Each lattice site may contain a defect which contributes to current transport. Voltage is applied to the OE whereas the AE is connected to GND. b) General simulation flow. In each increment of time, three steps are executed to calculate potential, current and a KMC transition.

by the current calculation and a KMC transition. These three steps and the relevant equations will be covered in the following sections.

2.4.1 Electric Field

To determine the electric field at each lattice site, a discretized version of the three dimensional Poisson equation is solved, reading

$$\Delta\Phi = -\frac{\rho}{\epsilon_0 \epsilon_\mathrm{r}}. \tag{2.5}$$

Here, Φ is the electrical potential, ρ is the charge density, ϵ_0 is the vacuum permittivity and $\epsilon_\mathrm{r} = 30$ is the relative permittivity of (tetragonal/cubic) HfO_2 [6]. With the potential, the electric field F can be calculated via

$$F = -\nabla\Phi. \tag{2.6}$$

2.4.2 Current and Temperature

The central part of the KMC model is the determination of the current through the defect network. Here, two conduction mechanisms are implemented, i.e. drift-diffusion (DD) and trap-assisted tunneling (TAT) [6]. Within a resistive switching cycle, the simulation chooses between both mechanisms in dependence of the number of oxygen vacancies in the oxide. Above a threshold number, the current is calculated regarding DD, otherwise using a TAT solver.

TAT

With the low density of defects in the HRS, it is assumed that no coupling between neighboring defects occurs [6] and thus, TAT is considered. Here, the current is composed by

(i) tunneling from electrode to trap,

(ii) hopping from trap to trap, and

(iii) tunneling from trap to electrode [3].

For the transitions (i) and (iii) between trap and electrodes, a hopping rate P is defined as

$$P = P_0 T_\mathrm{r} f, \tag{2.7}$$

with a coupling factor P_0. f represents the Fermi-factor which describes the distribution of occupied or free states in the metal electrode from or to which electrons may tunnel. The transition probability T_r is calculated using the Wentzel-Kramers-Brillouin (WKB) approximation [4, 181], reading

$$T_\mathrm{r} = \exp\left(-2\int_{x_\mathrm{t}}^{x_\mathrm{e}} \frac{1}{\hbar}\sqrt{2m^\star(e\Phi_\mathrm{B} - E_\mathrm{e})}\mathrm{d}x\right). \tag{2.8}$$

Here, x_e and x_t are the positions of electrode and trap. $m^\star$ is the effective tunneling mass, Φ_B is the tunneling barrier and E_e is the energy of the electron.
The electron hopping from trap to trap is implemented by calculating the electron occupational probability p_i for all defects. This is implemented by solution of a special form of the continuity equation [4]. This so-called "master equation" reads

$$(1-p_\mathrm{i})\sum_{j=1,j\neq i}^{N} p_\mathrm{j} h_\mathrm{ij} - p_\mathrm{i}\sum_{j=1,j\neq i}^{N}(1-p_\mathrm{j})h_\mathrm{ij} + (P_\mathrm{Ci}+P_\mathrm{Ai})(1-p_\mathrm{i}) - (P_\mathrm{iC}+P_\mathrm{iA})p_\mathrm{i} = 0, \tag{2.9}$$

and is solved by the Newton-Raphson iteration scheme [6]. The final current I_TAT is calculated with respect to one of the electrodes. Due to continuity the current at both electrodes has to be equal [6]. Exemplarily, the current at the anode is given by

$$I_\mathrm{TAT} = e\sum_{i=1}^{N}[p_\mathrm{i}P_\mathrm{iA} - (1-p_\mathrm{i})P_\mathrm{Ai}]. \tag{2.10}$$

DD

In the case of DD, the current density j is calculated self-consistently with the Poisson equation (2.5) by

$$\nabla \cdot j = 0. \tag{2.11}$$

From equation (2.11) the electron density n in the oxide is determined with the boundary condition

$$n_{\text{inter,i}} = n_0 T_{\text{r}}, \tag{2.12}$$

which connects the transition probability T_{r} from equation (2.8) to the electron densities at the oxide-electrode interface $n_{\text{inter,i}}$ and in the electrode n_0 [6]. Finally, the drift-diffusion current is given by

$$j = -e\mu_{\text{n}}(n\nabla\Phi - V_{\text{T}}\nabla n), \tag{2.13}$$

with the electron mobility μ_{n} and the thermal voltage V_{T}, given by

$$V_{\text{T}} = \frac{k_{\text{B}}T}{e}. \tag{2.14}$$

It may be noted here that in this dissertation, the DD mechanism is not used. All current calculations in the statistic models are performed with the TAT solver.

Temperature

To determine the temperature T at each lattice site, Fourier´s heat flow equation is solved [6], reading

$$\nabla \cdot (k_{\text{th}}\nabla T) = -g. \tag{2.15}$$

Here, k_{th} is the thermal conductivity of the oxide (0.5 $\frac{\text{W}}{\text{Km}}$ for HfO_x [4]) and g is the rate of heat generation, calculated by Joule heating:

$$g = F \cdot j. \tag{2.16}$$

2.4.3 Transitions

In the last step of the simulation flow (c.f. 2.6, b)), kinetic Monte Carlo transitions are selected and executed. Here, three possible processes are considered: Generation, recombination and diffusion. Based on the current state of the system determined in the previous steps, the rates of the three transitions are determined for all defects. In contrast to most published ReRAM KMC models [175–178], generation and recombination of oxygen vacancies is only allowed at the oxide electrode interfaces. This is

in line with the findings of Guo et al. [7] and Schie et al. [182] that the generation of anti-Frenkel pairs in the bulk oxide is unlikely.
The transition rates for generation R_G and recombination R_R are calculated by

$$R_G = \nu_0 \exp\left(-\frac{E_G - \alpha a e F}{k_B T}\right) \tag{2.17}$$

and

$$R_R = \nu_0 \exp\left(-\frac{E_R - (1-\alpha) a e F}{k_B T}\right). \tag{2.18}$$

Here, ν_0 is the attempt frequency, given by the characteristic lattice vibration. E_G and E_R are the respective energy barriers for generation and recombination. α represents a symmetry factor, a is the lattice constant and F is the locally induced electric field [6]. Analogous to the Arrhenius equation (2.2), the transition rate for the diffusion of oxygen vacancies to a neighboring lattice site is given by

$$R_D = \nu_0 \exp\left(-\frac{E_D}{k_B T}\right), \tag{2.19}$$

with the activation energy E_D.
After calculation of R_G, R_R and R_D for all lattice sites, one process is chosen randomly with probabilities weighted for the respective transition rates [6]. Subsequently, this single process is executed which represents the Monte Carlo step. Additionally, the transition rates are used to calculate the time increment Δt per simulation cycle via

$$\Delta t = \frac{-\ln(r_1)}{R_{\text{total}}}, \tag{2.20}$$

with a uniformly distributed random number r_1 drawn from the interval $[0, 1]$ and the sum of all transition rates R_{total} [183].

2.5 Normal and Log-Normal Statistics

The focus of this dissertation lies on the statistical evaluation of VCM ReRAM devices. This already implies the necessity to plot measured quantities for large sets of individual devices. One way to illustrate for example all measured cell resistances or read currents of a memory array is to plot the statistical distribution of the quantity. As will be demonstrated later, the data in this work is usually well described by a normal- or log-normal distribution. Therefore, this section outlines the generation of the respective plots and how they are to be interpreted.
A schematic of the probability density function (PDF) of a normal distribution is

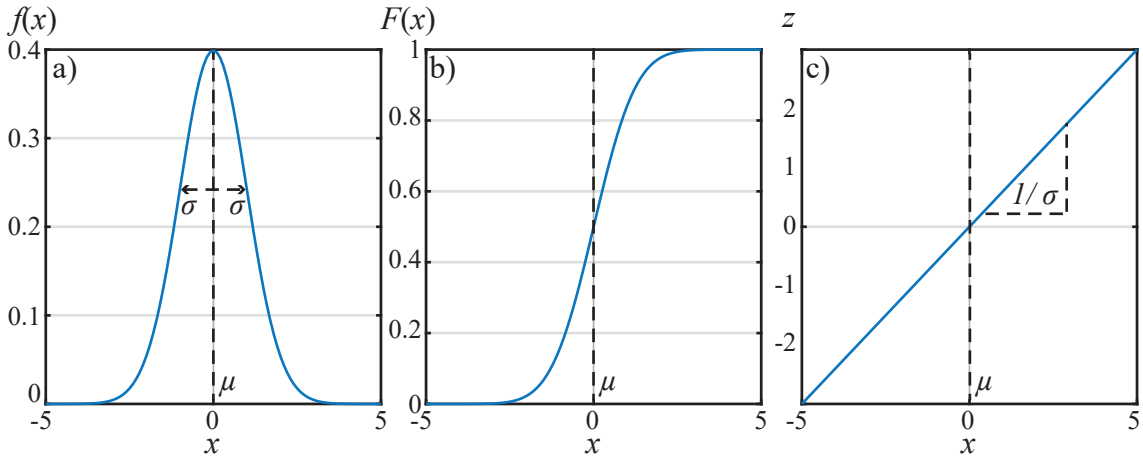

FIGURE 2.7: a) Schematic PDF of a normal distribution. b) Schematic CDF of a normal distribution. c) CDF normalized with regard to the standard normal distribution Φ. Here, z represents multiples of the standard deviation σ.

shown in figure 2.7, a). The typical Gaussian "bell" curve illustrates the probability $f(x)$ to observe a certain value of the quantity x. The highest probability is observed for the median μ of the distribution. Within $\pm$ one standard deviation σ lies 68.2 % of the data [184]. The probability density can be written as

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}. \tag{2.21}$$

However, throughout this work, the cumulative distribution function (CDF) is preferred which is given by the integral of the PDF, reading

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} \mathrm{d}t. \tag{2.22}$$

By using the Gauss error function (erf), equation (2.22) can be simplified to

$$F(x) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{x-\mu}{\sigma\sqrt{2\pi}}\right)\right). \tag{2.23}$$

A schematic CDF of a normal distribution is shown in figure 2.7, b). The cumulative probability $F(x)$, ranging from 0 to 1 (or 100 %), is read as the fraction of data points below or equal the respective value of x. Accordingly, 0.5 (or 50 %) represents the median μ. Though, with respect to the reliability of a memory device, usually the focus is not on the median of the distribution, but on the few worst bits in the tails of the distribution. As can be seen from figure 2.7, b), the value of the observed quantity x close to the edges of the distribution is difficult to read. This issue can be resolved by normalizing the CDF with respect to the standard normal distribution Φ which is

defined as

$$\Phi(z) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{z}{\sqrt{2\pi}}\right)\right). \tag{2.24}$$

Using the substitution

$$z = \frac{x - \mu}{\sigma}, \tag{2.25}$$

the normal distribution (2.23) can be written as standard normal distribution, reading

$$F(x) = \Phi(z). \tag{2.26}$$

Finally, z is obtained by calculating the inverse standard normal distribution of the distribution function $F(x)$, via

$$\Phi^{-1}(F(x)) = z = \sqrt{2}\mathrm{erf}^{-1}(2\Phi(z) - 1). \tag{2.27}$$

By plotting z instead of $F(x)$, the "S" shaped curve in figure 2.7, b) is converted to a straight line as shown in figure 2.7, c). Here, the resolution at the critical edges of the distribution is significantly improved. The probability axis is no longer ranging from 0 to 100 % but scales with multiples of the standard deviation.
Furthermore, this plot allows to identify normally distributed data right from the linear shape of the normalized CDF. According to the simple equation (2.25), the characteristic parameters μ and σ of the distribution can directly be read from the distribution, since μ equals $z(0)$ and σ is the reciprocal slope of the distribution. Additionally, the linearity can be utilized to extrapolate a limited data set easily towards larger statistics. For the mentioned reasons, the normalized CDF as shown in figure 2.7, c) is chosen as standard plot for most of the data throughout this work. It will be demonstrated later in this work that several data is not normally, but log-normally distributed. However, the log-normal distribution, reading

$$F(x) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{\ln(x) - \mu}{\sigma\sqrt{2\pi}}\right)\right), \tag{2.28}$$

only differs in the logarithmic dependence $\ln(x)$ from the normal distribution in equation (2.23). This means that the data is log-normally distributed if the logarithm of the observed quantity is normally distributed. Therefore, the same normalization is conducted, but the data is plotted with respect to $\ln(x)$. Analogously, a data set can be identified as log-normally distributed if a straight line appears on the normalized probability scale and a logarithmic scale for x.

3 Experimental Details

The experimental results in this work are obtained from two different material systems. On the one hand, comprehensive data for industrially fabricated VCM ReRAM devices based on HfO_2 were provided by *Infineon Technologies* and interpreted in this work. On the other hand, read variability, retention and endurance are analyzed for ZrO_2 based devices which are completely fabricated and characterized at *RWTH Aachen*. Since HfO_2 and ZrO_2 are nearly identical with respect to their physico-chemical properties [185], the experimental results for the two material systems are expected to be very comparable. This chapter covers the fabrication of these devices. Furthermore, the experimental setup developed in this work for the statistics based characterization of these cells is presented. Finally, typical experimental methods used to obtain the results in chapters 4 to 6 are explained.

3.1 Sample Fabrication

In order to investigate reliability aspects with respect to intrinsic statistics, the experimental characterization of single cells becomes insufficient. For an efficient characterization of many cells, a suitable sample structure (and test setup) is required. Therefore, array structures come into account, which can be distinguished between active and passive arrays [186]. Whereas ReRAM can easily be incorporated into existing CMOS structures to form memory cells consisting of one transistor and one resistor (1T1R) [187], such active array structures are rather complicated to fabricate from scratch. Additionally, the characteristics of the processed transistor would superimpose the investigated reliability aspects of the ReRAM cell. Therefore, only passive arrays come into account for this dissertation.
However, passive cross-bar structures are reported to suffer from parasitical sneak paths which scale with the size of the array [188]. Thus, a one-dimensional 32x1 sub-array is designed in this work which provides 32 individual cells connected to one common bottom electrode. The structure and fabrication is published in [189] (accepted) and explained in the following.
Figure 3.1, a) depicts a schematic of the fabricated cell stack and figure 3.1, c) shows a sketch of the array structure. The devices are deposited on a Si wafer covered by an

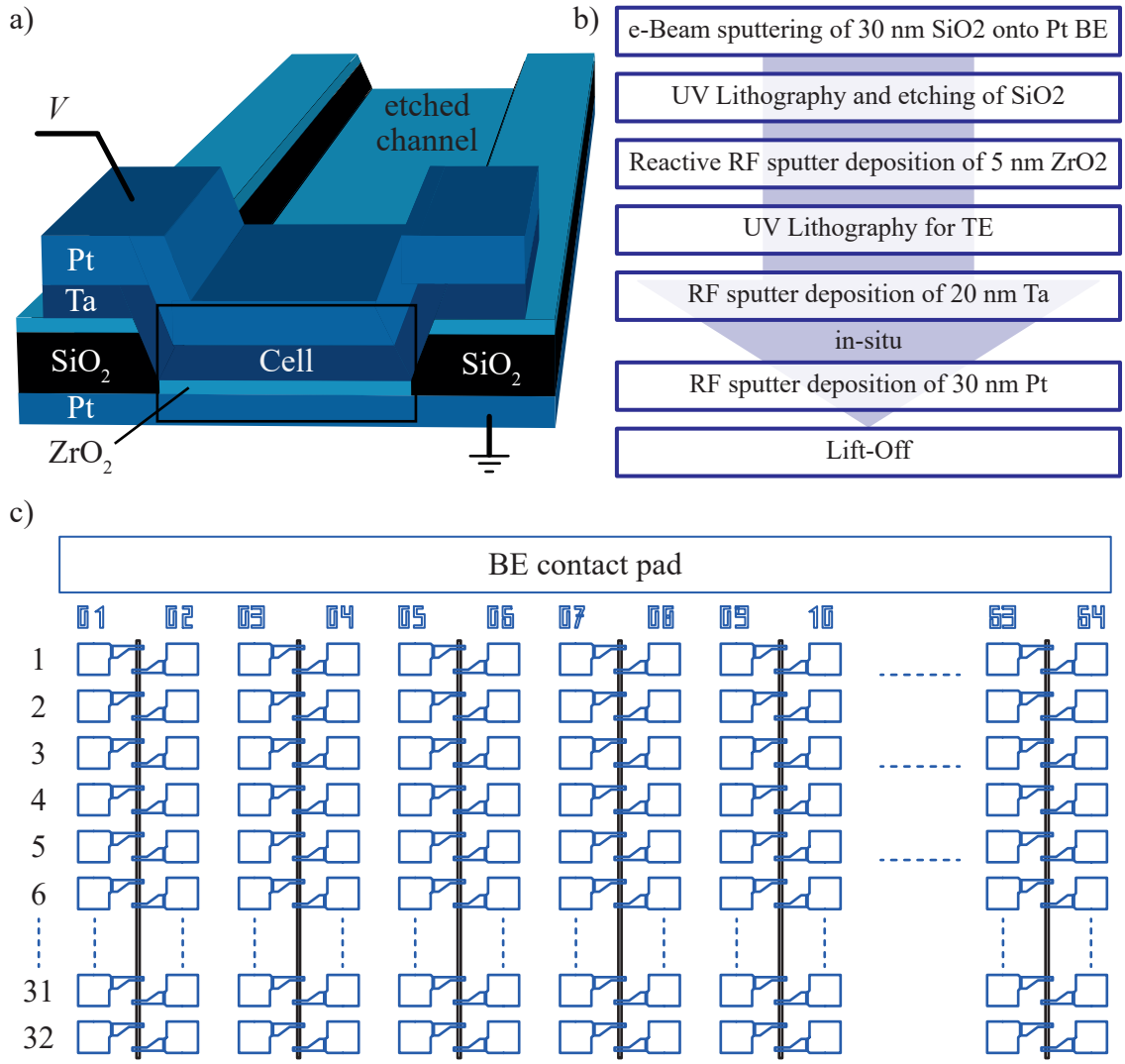

FIGURE 3.1: Sample fabrication. a) Schematic of the device stack. A Pt channel is etched into a SiO_2 layer and serves as bottom electrode. The cell is located at the cross section of this channel and the top electrode pad. b) Fabrication process flow diagram. c) Sketch of the array structure. For each cell size, the sample provides 64 individual sub-arrays with 32x1 cells each. Partially reproduced from [189].

SiO_2 insulation layer. Via RF sputtering, a 5 nm Ti layer is added, followed by 30 nm Pt. the latter will be the bottom electrode of the device and is shown as base layer in figure 3.1, a). The Ti layer is required to improve adhesion of Pt on SiO_2 (not shown). Subsequently, the Pt covered wafer is cut into 1 in x 1 in substrates.

As summarized by the flow diagram in figure 3.1, b), the Pt bottom electrode is covered completely by 30 nm e-Beam sputtered SiO_2. This layer is used to separate the single sub-arrays and is chosen sufficiently thick to not participate in resistive switching. Using UV lithography, the SiO_2 layer is structured according to the black vertical lines in figure 3.1, c). Subsequently the SiO_2 along these lines is etched down to the underlying Pt. this creates the SiO_2 free channel which contains the cell stack

in figure 3.1, a).
After removing the remaining photo-resist, 5 nm ZrO_2 are deposited via reactive RF sputter deposition onto the whole sample (SiO_2 and Pt channels). The ZrO_2 layer on top of SiO_2 has no function. However, within the Pt channels, ZrO_2 represents the resistive switching oxide layer.
Another UV lithography step follows in order to structure the top electrodes corresponding to the blue outlined pads in figure 3.1, c). Subsequently, 20 nm Ta are deposited as ohmic electrode via RF sputter deposition. To prevent oxidation of the Ta layer, it is in-situ covered by 30 nm RF sputtered Pt. It may be noted that Ta represents the standard ohmic electrode metal in this work, since ZrO_2/Ta devices are reported to have excellent resistive switching characteristics [190]. However, for the endurance characterization in chapter 6, the metal for this layer will be varied between Ta, Ti, Hf and Zr. A lift-off process removes the residual photo-resist, Ta and Pt outside the top electrode pads and finalizes the sample. As indicated in figure 3.1, a), the sample provides 64 individual sub-arrays with 32x1 cells each. Moreover, the sample includes four different cell sizes, being (3 µm x 3 µm), (5 µm x 5 µm), (7 µm x 7 µm) and (10 µm x 10 µm).
Microscopic pictures of the device structure are provided in figure 3.2. A light microscopy of the top view is shown in figure 3.2, a). Here, the brown area is the passive SiO_2 layer (covered by 5 nm ZrO_2). Along the vertical, white lines SiO_2 is etched down to Pt which is subsequently covered by the ZrO_2 switching layer. Perpendicular to these lines, the top electrode pads form cross-bar cells as highlighted by the black squares. At the edge of the sample, the common bottom electrode is accessible by a large pad which was covered during all deposition steps (not shown). Figure 3.2, b) shows a 7 µm x 7 µm cell captured by scanning electron microscopy (SEM). The horizontal top electrode covers the vertical bottom electrode in the center of the picture. The dark gray area is covered by SiO_2. The SEM image reveals a width of the electrode structures of approx. 9 µm which is 2 µm broader than the intended 7 µm. Figure 3.2, c) shows a detailed study of a 3 µm x 3 µm cell. Again, the real width of the electrodes is observed to be increased by 2 µm, resulting effectively in a 5 µm x 5 µm cell. The higher magnifications in the second and third frame (from left to right) show a rather high roughness of the top-electrode edge of approx. 0.5 µm. Both the pronounced roughness and the increased width of the structure might be explained by the quality of the photo-resist used during the lithography steps. Presumably, the quality of the structure could be improved by optimization of the photo-resist composition and exposure. However, minor variations in cell area should not affect the resistive switching characteristics of filamentary VCM ReRAM, because the radius of the filament is expected to be significantly smaller than the cell.

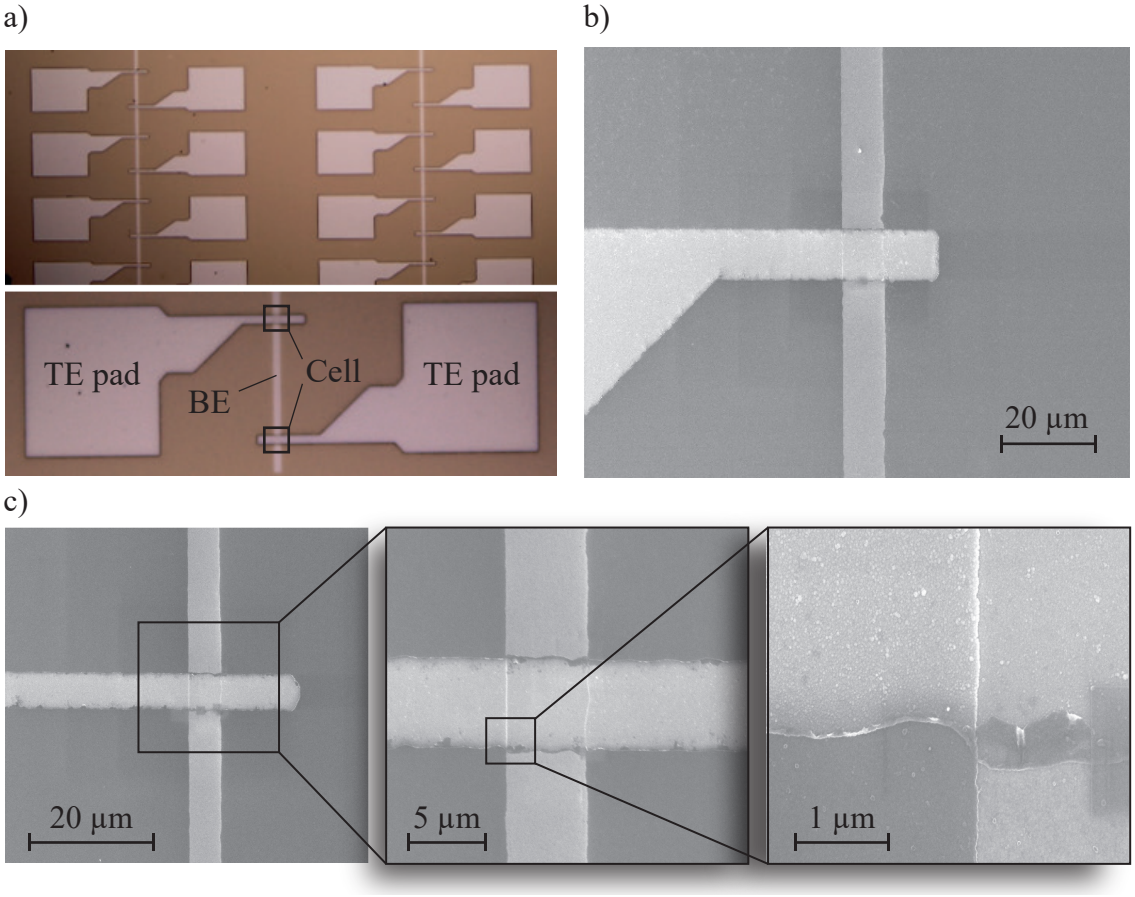

FIGURE 3.2: Microscopy of the fabricated (30 nm Pt / 5 nm ZrO_2 / 20 nm Ta / 30 nm Pt) cross-bar devices. a) Top-view light microscopy of the array structure. the vertical BE lines are etched in the brown SiO_2 layer. At the cross-sections with the horizontal TEs the switching cells are located. b) SEM image of a single (7 µm x 7 µm) cell. c) Detailed SEM study of a (3 µm x 3 µm) cell. In both cases, the real width of BE and TE is observed to be increased by 2 µm. The roughness of the TE edges is approx. 0.5 µm.

The advantage of this structure is the potential to contact 32 top electrodes by a suitable probe card (with 32 probes) while the common bottom electrode can be connected to ground, as will be outlined in the following section. However, one drawback compared to conventional cross-bar structures [191–193] is the capacity between the large top electrode pads and the common bottom electrode. By using a rather thick SiO_2 layer the impact of this capacity should be minimized. Nevertheless, it may affect the resistive switching characteristics and could cause parasitic current overshoots when the cell current drops during forming or SET operations.

3.2 Experimental Setup

In order to characterize the fabricated sub-array structure, a dedicated test setup is developed in this work. The sample is placed in a *SemiProbe LA-100* probe station

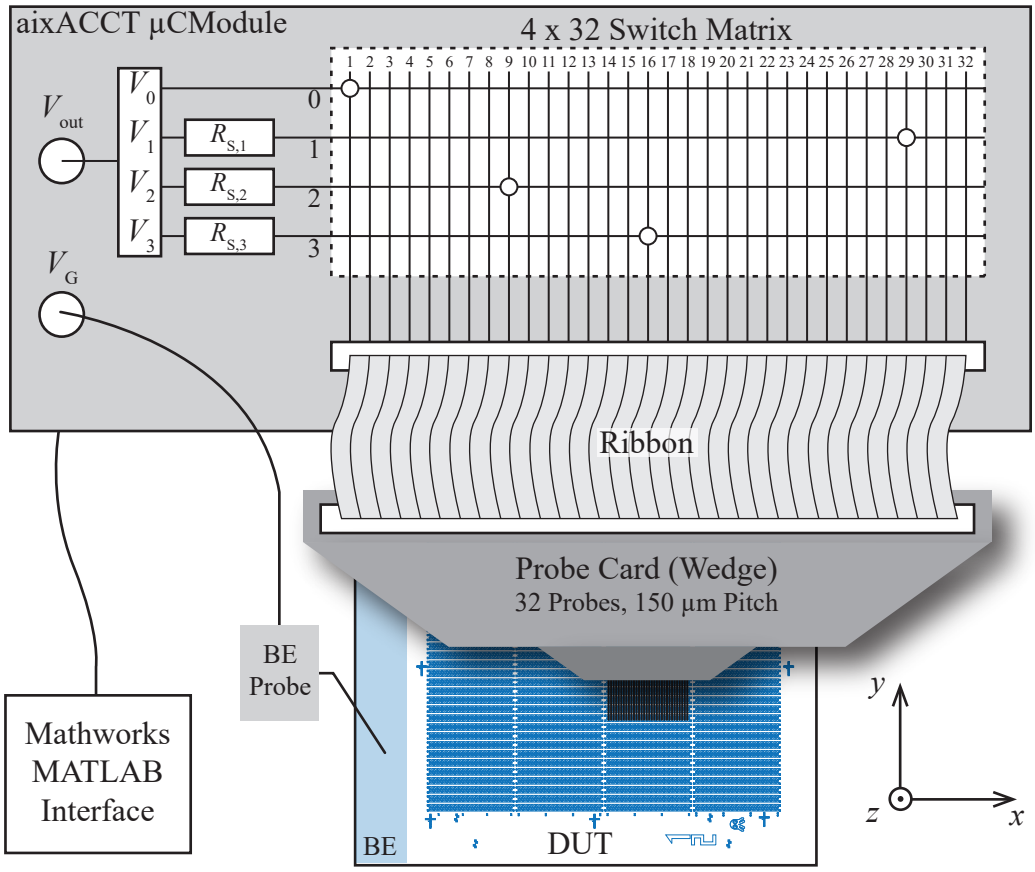

FIGURE 3.3: Schematic of the experimental setup developed in this work. The custom array tester based on the *µCModule* platform by *aixACCT Systems* provides a voltage output V_{out} with AWG DAC, mapped to 4 parallel voltage channels $V_{0\text{-}3}$. These channels are connected via optional series resistors $R_{S,1\text{-}3}$ to a 4x32 switch matrix. Here, the voltage signal can be mapped arbitrarily to 32 output channels which are connected via ribbon cable to a wedge probe card. By contacting the 32 probes of the card to the DUT, the desired voltage signal can be applied to the respective cells. The common BE is connected by an additional probe to the virtual ground V_G. In order to control the device and for data collection an interface to *Mathworks MATLAB* is developed.

which provides a vacuum chuck capable of planar rotation (in x-y-plane, c.f. figure 3.3). This degree of freedom is crucial to adjust the angle of the sample to the orientation of the probes.

As depicted by the schematic experimental setup in figure 3.3, the sample or (also *device under test* DUT) is contacted by a wedge probe card with 32 probes in one row. According to the array structure, the probes are spaced with a pitch of 150 µm. The probe card is mounted to a micro-positioner via a probe-arm which was originally designed for high frequency (HF) probe heads. This arm not only ensures a stable mounting of the probe card but also allows for fine tilting around the y-axis (c.f. figure 3.3) which is desirable to parallelize the array of probes to the sample. 32 probes with 150 µm pitch result in a probe array length of 4.65 mm which results

in significant differences in height if the probe card and sample are not perfectly parallel.
The probe card is electrically connected via a ribbon cable to a custom array test device developed by *aixACCT Systems* and based on their *μCModule* platform. It provides a digital analog converter (DAC) with arbitrary waveform generator (AWG) and a maximum sampling rate of 1 MHz. The emitted waveform V_{out} can be mapped to 4 voltage output channels $V_{0\text{-}3}$. Here, V_0 is equal to V_{out}, whereas $V_{1\text{-}3}$ can be adjusted to fractions of V_{out} by 3 independent potentiometers. Additionally, V_3 may optionally be inverted. However, in this work neither the potentiometers nor the inverter are used. $V_{0\text{-}3}$ are treated as four parallel sources of V_{out}. Instead of using the potentiometers, 3 different series resistors $R_{\text{S},1\text{-}3}$ are added in this work to $V_{1\text{-}3}$. For a typical switching operation, only one of the four channels will be used as voltage source. Here, the setup enables to choose between different series resistors depending of the desired operation. Whereas a RESET operation may be performed without series resistor, during forming and SET the voltage divider induced by a series resistor can be beneficial to limit the current during switching.
The most important component of the array tester is a 4x32 switch matrix. Here, the four voltage output channels (rows) are mapped to 32 output channels (columns). As indicated by the circles in figure 3.3, the rows and columns can be connected arbitrarily. In the shown example V_0 is connected to column 1, V_1 via $R_{\text{S},1}$ to column 29, V_2 via $R_{\text{S},2}$ to column 9 and V_3 via $R_{\text{S},3}$ to column 16. The 32 output channels (columns) are provided by a 32x2 pin ribbon connector with 32 signal lines and 32 shielding lines which are connected to ground. The 64 lines of the attached ribbon cable carry signal and ground in an alternating order to prevent crosstalk between signal lines. Via this ribbon cable, the 32 output channels are connected to the wedge probe card with 32 probes towards the sub-array sample structure.
The common bottom electrode (BE) of the sample is connected via an additional probe to the virtual ground V_{G} input of the array tester. This input serves as ground reference for the sample and provides an analog digital converter (ADC) with a maximum sampling rate of 0.8 MHz. Here, the current through the sample is recorded.
Based on the command library (dll) provided by *aixACCT Systems*, an interface to the software *Mathworks MATLAB* is developed in this work. Here, the *MATLAB MEX* environment is used to implement *MATLAB* functions compiled in C++ which access the device functions provided by the according dll. These *MATLAB* functions enable to access the switch matrix in order to set the desired connections between the input rows (0-3) and the output lines (1-32). Subsequently, a measurement can be started by sending an arbitrary waveform (and additional parameters) to the measurement device. For a more efficient and user friendly operation, the basic functions are

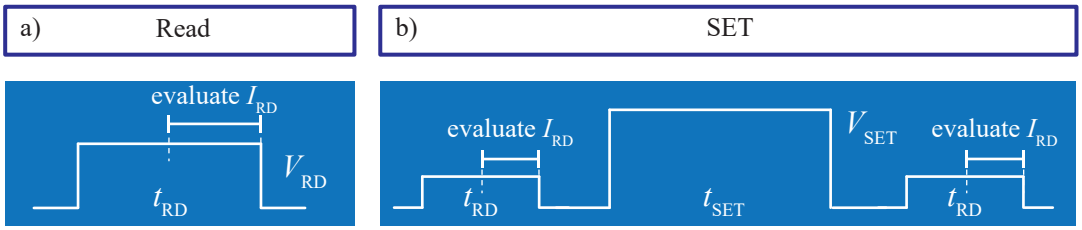

FIGURE 3.4: Basic operations with the array test setup. a) The standard read function applies a rectangular voltage pulse with height V_{RD} and width t_{RD}. The function returns the resulting current as raw data as well as the current median of the second half of the pulse. b) A basic SET operation reads the state of the cell before and after the programming pulse with height V_{SET} and width t_{SET}. Both functions can be applied to multiple cells and return the evaluated current states of the respective cells.

wrapped into higher level *MATLAB* functions for distinct operations, like SET, RESET or Read.

The developed *MATLAB* interface provides a script based environment for controlling the array tester as well as data collection. Due to the computing capabilities of *MATLAB*, data can be evaluated within the run-time of a measurement. This enables for example the manipulation of the switching parameters based on the results of the preceding switching or reading operation. In particular, it allows for switching voltage optimization during endurance experiments and program-verify algorithms as explained in the following section.

3.3 Experimental Methods

3.3.1 Basic Operations

As explained above, the *aixACCT Systems* array tester is controlled via an interface to the software *Mathworks MATLAB*. Here, several functions are implemented to perform certain operations automatically. The most simple operation is to read the current state of a contacted cell. In order to do this, a "Read" function is provided which applies a rectangular voltage pulse as sketched in figure 3.4, a), with an arbitrary pulse length t_{RD} and height V_{RD}. The resulting current signal measured by the device is returned as raw data into a *MATLAB* array. Additionally, the read current is evaluated directly by calculating the current median of the second half of the pulse. Since the measurement setup is designed to characterize a 32x1 sub-array of cells, the function can be called for an arbitrary set of one to 32 cells which will be read consecutively. Besides the raw data, the function returns an array with the evaluated read current state of each cell. Thus, the setup enables a fast assessment of the current state of up to 32 cells.

Likewise, an arbitrary set of cells within one 32x1 sub-array can be programmed by a dedicated SET or RESET function. The former is schematically shown in figure 3.4, b). Typically, a single programming sequence consists of a programming (e.g. SET) pulse with arbitrary width (t_{SET}) and height (V_{SET}), flanked by two read pulses. This means that the state of the cell is read before and after programming. The function returns the raw data as well as the evaluated states before and after the programming pulse. If multiple cells are programmed, an array of current states is returned. A RESET is performed analogous to the SET scheme in figure 3.4, b) by reversing the polarity of the programming pulse. By increasing the pulse height and length, the SET function can also be used to electroform cells. However, a more reliable electroforming operation is achieved by applying a triangular instead of a rectangular pulse.
The four individual voltage input rows (0-3) can here be used to select an appropriate series resistance for each operation. For the standard devices in this work, usually $R_{\mathrm{S}} = 10\ \mathrm{k}\Omega$ is selected for electroforming and $R_{\mathrm{S}} = 1\ \mathrm{k}\Omega$ for SET operations. The RESET is performed without external series resistance.
In conclusion, the setup can be used to conveniently read and program the states of 32 contacted cells. Each programming operation directly returns the current state of each cell before and after programming which allows for a fast assessment of the programming success.

3.3.2 Program-Verify

Since the success of each programming operation is directly monitored, unsuccessful operations can be corrected on the fly. In order to do this, program-verify algorithms come into account [89]. Figure 3.5 shows flow charts of program-verify algorithms used in this work. A typical forming algorithm performs a triangular voltage pulse with a slope time of 20 ms to reach the stop voltage V_{stop}, followed by a read pulse as introduced above. The resistance of the cells is evaluated after the forming pulse and compared to the target resistance R_{target}. A cell resistance below this target is considered as successful forming. Each cell with a resistance $R > R_{\mathrm{target}}$ will receive another forming pulse with increased stop voltage. As soon as all cells formed successfully or the maximum number of iterations is exceeded, the algorithm is terminated and the resulting states of the cells are returned. Optionally, all unsuccessful cells after the maximum number of iterations can be discarded for any further operations. Thus, a single function is executed for a set of 32 virgin cells which returns a set of successfully formed cells with the according resistances before and after forming.
Subsequently, the successful cells can be cycled using the RESET and SET algorithms

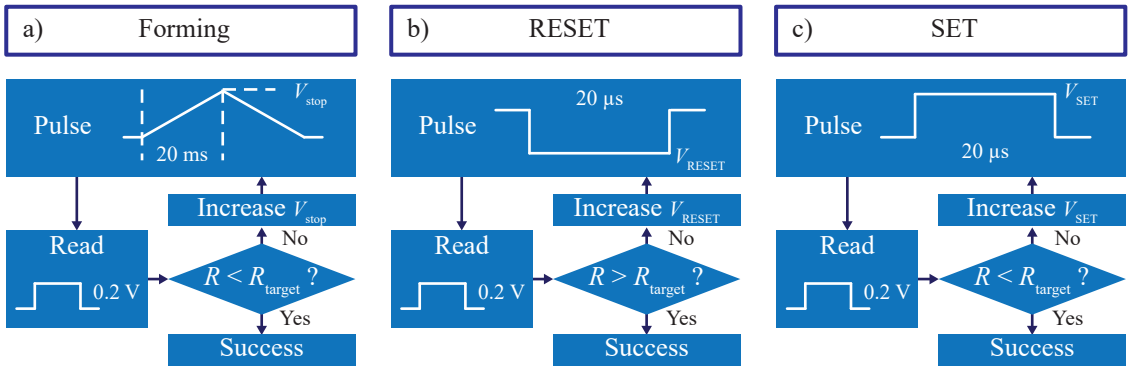

FIGURE 3.5: Schematic of the program-verify algorithms. a) The forming algorithm applies a triangular voltage pulse with a rise time of 20 ms towards the stop voltage V_{stop}. After each programming pulse, all cells are read and their resistances compared to the target resistance R_{target}. In case of failed programming, the pulse is repeated with increased stop voltage until all cells passed or the maximum number of iterations is reached. The RESET and SET algorithms in b) and c) are performed analogously using rectangular programming pulses with a width of 20 µs and the height V_{RESET} or V_{SET}, respectively. Adapted from [189].

in figure 3.5, b) and c). Here, typically rectangular voltage pulses with a width of 20 µs are applied. Analogous to the forming algorithm, the resistance after the SET or RESET pulse is evaluated and compared to a target resistance R_{target}. Again, all failed cells are programmed again with increased pulse height V_{SET} or V_{RESET} until all cells are successfully programmed or the maximum number of iterations is reached.
In conclusion, the implemented program-verify methods provide a highly automated way to electroform and cycle multiple VCM ReRAM cells. Defect cells are discarded automatically and functioning cells can be programmed into defined resistance (or read current) margins with a comparatively low stress, since the programming parameters are initiated low and only increased if the initial parameter was insufficient. The ability to program cells into distinct margins is utilized in [189].

3.3.3 Endurance

The discussed program-verify algorithms are well suited to conveniently program up to 32 cells into defined margins. However, in order to assess the endurance of the devices under test, SET/RESET cycle numbers of 10^6 or higher have to be performed within reasonable time scales. Here, the presented algorithms are too slow because every programming operation is performed separately. Each call of the low level measurement function along with data transfer between measurement device and computer takes approx. 1 s. Assuming that every programming pulse succeeds, one cycle (SET + RESET) would cost approx. 2 s. Thus, an endurance measurement over 10^6 cycles would require $2 \cdot 10^6$ s or more than 23 days.

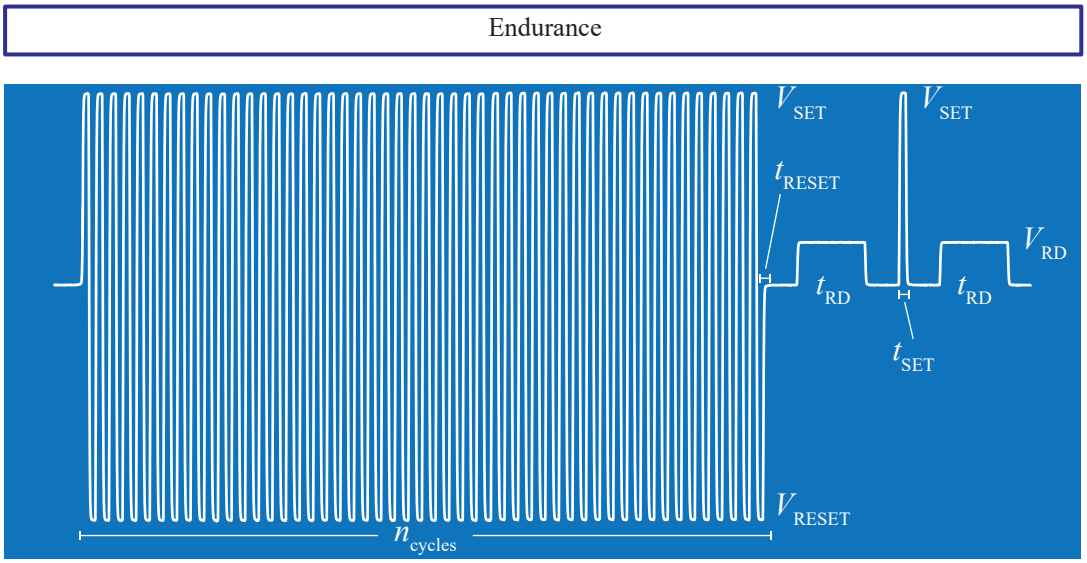

FIGURE 3.6: Schematic of the endurance pulse sequence. To achieve high numbers of cycles within reasonable measurement time, n_{cycles} cycles are performed consecutively without read pulse in between. Each cycle consists of one SET pulse with t_{SET} and V_{SET} and one RESET pulse with t_{RESET} and V_{RESET}. During the last cycle of each sequence, one HRS and one LRS is measured and evaluated.

It is therefore desirable to apply as many cycles as possible within one call of the low level measurement function (one executed waveform and data transfer). Hence, an endurance function is implemented which performs a high number n_{cycles} of SET/RESET cycles in a row, as depicted schematically in figure 3.6. The waveform comprises n_{cycles} consecutive SET and RESET pulses without reading in between. In order to determine if the tested cell switches sufficiently, read pulses are added to the last cycle of the sequence. Thus, every n_{cycles} cycles, the HRS and LRS of the cell is evaluated once. Depending on the result, the programming parameters can be adjusted for the following sequence.

This induces a trade-off between cycling speed and the number of possible adjustments. The more cycles are performed in one sequence, the faster the measurement can be completed, but the less reads are performed which allow for manipulation of the cycling parameters. With the available buffer size of the measurement device (approx. 22000 samples), a maximum of 10^3 cycles can be performed in one sequence, if the pulse length is $t_{SET} = t_{RESET} = 10$ µs. Thus, 10^6 cycles can be performed in approx. 17 min. To find the best balance between speed and number of reads, an endurance algorithm is developed in this work which adaptively changes the number of cycles per sequence according to the success rate of the current switching parameters. Starting with a low number of cycles per sequence, this number is increased each time the switching parameters did not need adjustment. The developed algorithm is discussed in detail in chapter 6.

In general, the measurement device enables to apply voltage to multiple cells in parallel. Since only one current can be measured, reading the exact state of multiple cells in parallel is not possible. However, multiple cells could be cycled in parallel and read consecutively after n_{cycles}. Although this seems like a faster option to determine the endurance of multiple cells, it actually provides no advantage over measuring endurance one cell after the other. The reason is that, as mentioned above, the number of function calls limits the effective speed. As stated above, 10^6 cycles on a single cell can be performed by 10^3 sequences containing 10^3 cycles each. This results in 10^3 s or approx. 17 min. Respectively, 10 cells measured consecutively would cost $10 \cdot 17$ min$= 170$ min.
If these cells were cycled in parallel, this would cost only 17 min. However, after each sequence of 10^3 cycles all cells would need to be read separately. Thus, 11 function calls (cycle + 10 reads) would be called 10^3 times, resulting in approx. 183 min. Thus, it is faster to assess cell by cell and include the read operation into the cycle function. Nevertheless, the setup allows to automate the endurance measurement of all 32 contacted cells. It therefore enables to generate reasonable statistics within justifiably measurement times.

3.4 Conclusion

Within the scope of this work, an experimental environment was developed which enables the efficient characterization of multiple VCM ReRAM cells in order to generate statistically sound data within reasonable time scales. A sample structure providing 32x1 sub-arrays of cells was designed and tested successfully. A respective test setup including a suitable probe card was built around the array tester provided by *aixACCT Systems*. The development of an interface between the array tester and *Mathworks MATLAB* allowed for implementation of several measurement routines with focus on automation and statistical evaluation. This includes program-verify algorithms for the reliable programming into defined resistance (or read current) margins, as well as comparatively fast and automated endurance routines. The measurement results obtained with the presented experimental environment will be presented in the following chapters. For the experimental characterization of variability (c.f. chapter 4) and retention (c.f. chapter 5), the implemented program-verify and read operations are of high importance. The fast cycling methods are utilized for endurance measurements as presented in chapter 6.

4 Variability

The pronounced stochastic nature of VCM ReRAM is one of the largest challenges for its industrial adaption. Apart from fabrication induced device to device (D2D) variations, the programmed states fluctuate significantly from cycle to cycle (C2C) and even from read to read (R2R). This chapter briefly covers D2D and C2C variability. Subsequently, R2R fluctuations are identified as key challenge regarding the variability of VCM ReRAM and studied in detail. Therefore, this chapter provides experimental observations of read variability and two different model approaches towards the understanding of its physical origins.

4.1 Device to Device and Cycle to Cycle Variability

As reported in [86, 87], the cycle to cycle fluctuations in typical VCM ReRAM are expected to exceed the stochastic differences between individual devices. Therefore, D2D and C2C variability are reported to be indistinguishable [86, 87]. In order to verify this observation for the standard (30 nm Pt / 5 nm ZrO_2 / 20 nm Ta / 30 nm Pt) cross-bar devices investigated in this work (c.f. chapter 3.1), a comparison of D2D and C2C variability is given in figure 4.1. The lines in this plot represent the distributions of HRS and LRS resistance drawn from an endurance experiment (c.f. chapter 6). 51 individual cells of the standard stack mentioned above are cycled continuously until failure which occurs as soon as the cell becomes irreversibly stuck in either LRS or HRS. Each programming pulse has a width of 20 µs. By default, the SET voltage is 0.9 V and the RESET voltage -1.7 V which are both applied to the Ta electrode with Pt connected to ground. However, as will be discussed in chapter 6, the programming voltages are adaptive between cycles to achieve the highest possible endurance.
The resulting LRS and HRS resistance distributions show significant C2C fluctuations (blue lines). This also includes few cycles with unsuccessful SET or RESET respectively. For comparison, a distribution for a single cycle over all 51 devices is depicted by the black dashed line. As expected from [86, 87], the C2C and D2D distributions are rather similar.
Additionally, the impact of a program verify algorithm on D2D and C2C variability is studied. The orange symbols represent D2D distributions of > 200 cells drawn

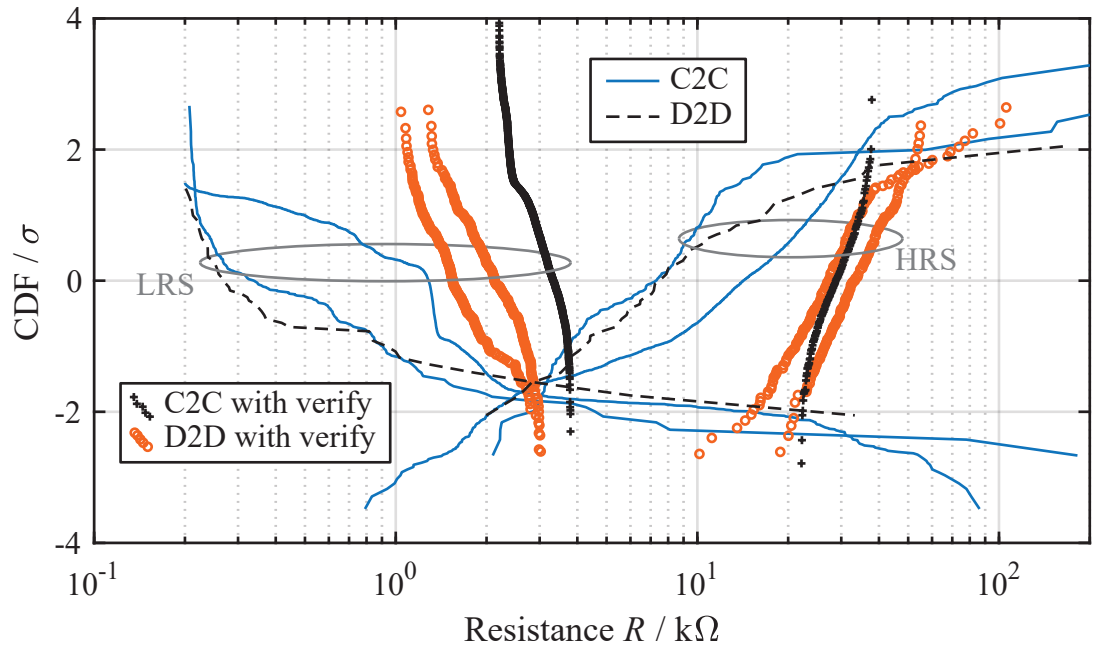

FIGURE 4.1: Comparison of C2C and D2D variability. The lines are drawn from an endurance experiment. The distribution of one cycle over 51 individual cells (black dashed) is comparatively similar to C2C distributions for single cells (blue lines). The symbols represent C2C (black crosses) and D2D (orange circles) distributions drawn from another experiment. Here, a program-verify algorithm is used which significantly reduces the variability. Again C2C and D2D distributions are rather similar.

from another experiment which are programmed by the program-verify algorithm introduced in chapter 3.3. Using this algorithm, the cells are progressively exposed to pulses of increasing amplitude until a defined threshold resistance is exceeded. This leads to much steeper distributions and thus reduces the observed D2D variability significantly. For comparison, the black cross symbols represent a C2C distribution measured on a single cell. It can be seen that the C2C variability highly benefits from the program-verify approach as well. The resulting distribution is again very similar to the orange D2D distributions.
It is therefore concluded that C2C and D2D variability are indistinguishable in the investigated ZrO_2 based devices. It can furthermore be reduced to large extend by application of a suitable program-verify algorithm.

4.2 Read to Read Variability

As demonstrated above and reported in [89], both D2D and C2C variability become a secondary issue since it can be tackled by appropriate programming algorithms. In contrast to this, R2R variability seems to be an intrinsic characteristic of filamentary VCM ReRAM. Stochastic changes of the programmed state from read to read limit

the effectiveness of programming algorithms [86, 103] and determine the maximum read window between HRS and LRS. Therefore, this type of variability is identified as key challenge and investigated in detail over the following sections. These will cover experimental results as well as two different approaches to model and understand R2R variability.

4.2.1 Experimental Results

At first, experimentally obtained results are presented for two different material systems. The first subsection covers read variability in the ZrO_2 based devices. Subsequently, the results are extended by the industrially fabricated devices based on HfO_2.

ZrO_2

The read variability is investigated using the standard ZrO_2 cross-bar devices discussed in chapter 3.1. Using the test setup presented in chapter 3.2, 20 individual cells are electroformed and undergo 10 initial switching cycles. Here, all forming and switching operations are performed using a program-verify algorithm as introduced in chapter 3.3. The forming voltage is initialized at 1.8 V and is increased up to a maximum of 4 V if necessary. The set voltage ranges from 0.6 V to 1.4 V and the reset voltage from 1.3 V to 2 V. After each programming pulse, the cell resistance is verified by a read pulse of 0.35 V for 20 µs.
To determine the read variability, each cell is read by applying the read voltage of 0.35 V constantly for a duration of 200 ms. The resulting current is recorded with a sampling rate of 80 kHz. This read is repeated 10 times before concatenating the current traces to a total length of 2 s as shown in figure 4.2. After the read sequence of 10 reads is completed, the cell is cycled again and subsequently the read sequence is repeated. By repeatedly alternating switching and reading, 540 HRS states are recorded. As reported in [86, 87] and demonstrated above, C2C variability is indistinguishable from D2D variability. Therefore, the recorded states are considered equal to 540 individual cells. To ensure comparability, only cells with a read current between 6 µA and 12 µA are included.
Within the first 600 ms of the current traces, marked as region I in figure 4.2, the states are observed to be very unstable and comprise large fluctuations. Here, the read current seems to drift towards more stable values. Subsequently, in region II, the traces are characterized rather by fluctuations around a relatively stable current median.
To demonstrate the impact of read noise on the reliability of VCM ReRAM as memory

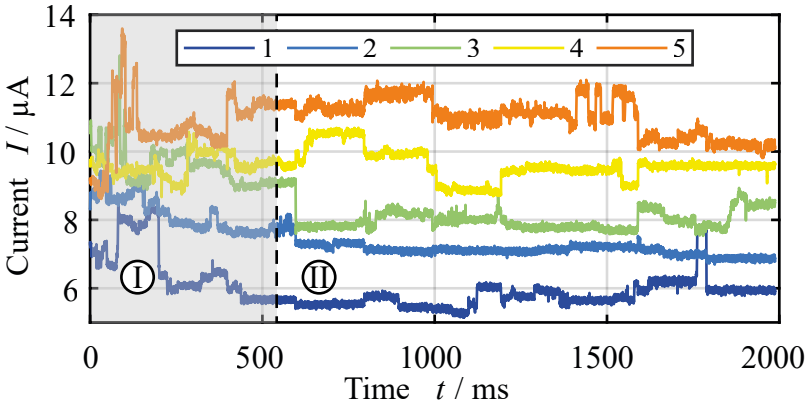

FIGURE 4.2: Read current traces of five exemplary (30 nm Pt / 5 nm ZrO_2 / 20 nm Ta / 30 nm Pt) cells in HRS. The traces are concatenated from 10 reads at 0.35 V for 200 ms each. The data is divided into two regions. (I) from 0 ms to 600 ms characterized by large fluctuations and drifting of the current median. (II) 600 ms to 2 s comprising random fluctuations around a largely constant current median. Reproduced with permission from S. Wiefels et al. [194], © 2020 IEEE.

device, the data is plotted as read current distribution in figure 4.3. Here, each line is a cumulative read current distribution of all cells at one instant of time. As mentioned above, only cells in the range of 6 µA to 12 µA are included in the data set. This filter is applied at the time $t = 12.5$ ms. Thus, the respective distribution in figure 4.3, a) is constraint to this interval. At later instants of time, the read current exceeds these limits and the distribution relaxes to a straight line. With the normalized probability scale (σ scale) and the logarithmic current scale, a linear shape identifies the data as log-normally distributed which is common for the HRS [123, 124]. Additionally, within the first 600 ms, the distribution tilts to a shallower slope. This broadening can be accounted to the significant instability observed in this region (I). After 625 ms, the distribution, including slope and log-normal shape is observed to be very stable. With the purpose of ensuring a sufficiently large read window, distribution shaping is a common approach [195]. The distribution is shaped by either reprogramming or discarding all cells which exceed a defined threshold current or resistance. This concept is applied to the stable region II in figure 4.3, b). Here, all cells with a read current above 12 µA at $t = 750$ ms are removed from the data set. For $t = 800$ ms up to $t = 1500$ ms only the remaining cells are considered when plotting the read current distribution. In theory, no cells should exceed the limit of 12 µA as observed for the shaped distribution at 750 ms. However, within another 750 ms (towards $t = 1500$ ms) the removed "tail" recovers and the log-normal shape of the distribution is restored. This means that the effect of shaping algorithms is volatile and their potential to widen the read window is limited.

Without (data-)shaping it is remarkable that the read current distribution in region II

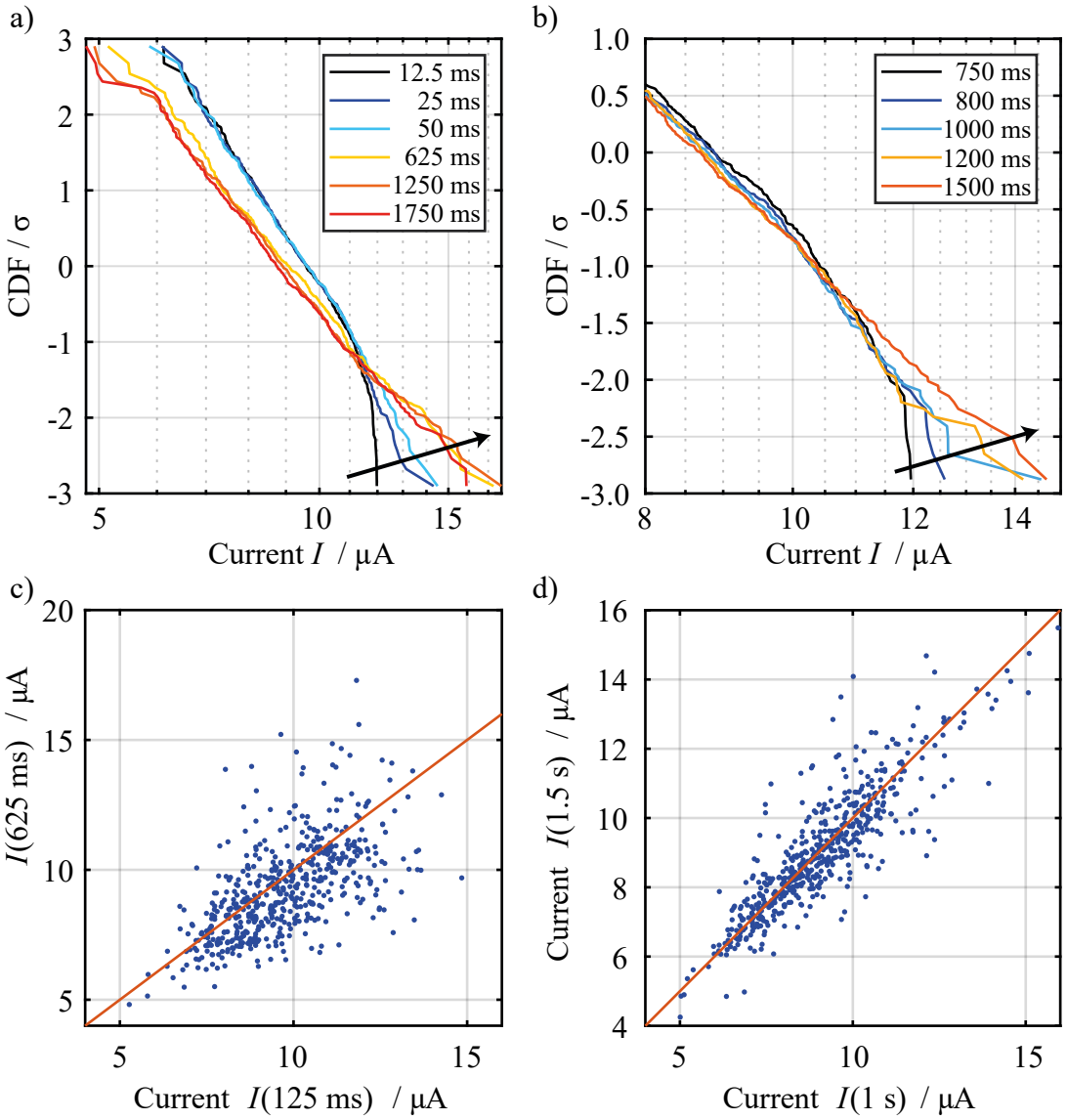

FIGURE 4.3: a) Read current distributions of 540 HRS states for different times corresponding to the current traces in figure 4.2. Only states between 6 µA and 12 µA are included, filtered at 12.5 ms. The distribution relaxes to a log-normal shape and broadens until 625 ms. Subsequently, the log-normal distribution is stable. b) Data shaping experiment in the more stable region II. At 750 ms all cells above 12 µA are discarded. The remaining cells are plotted for 800 ms to 1500 ms and the log-normal shape recovers without the removed cells. c) Lag plot showing current changes (blue dots) within 500 ms in region I. The red line represents no current change. d) Analogous lag plot for region II. Partially reproduced with permission from S. Wiefels et al. [194], © 2020 IEEE.

is highly stable. The yellow, orange and red lines in figure 4.3, a) coincide to a great extend although the individual current traces comprise significant fluctuations around their median (c.f. figure 4.2). These fluctuations can also be depicted by lag-plots as shown in figure 4.3, c) and d). For all 540 cells, the read current at one incident of time is compared to a later one (blue dots). The deviation from the red diagonal represents the current change between the two reads. figure 4.3, c) compares the read current at 125 ms and 625 ms, in the unstable region I. As expected, large fluctuations of up to 6 µA within the considered 500 ms are observed. Only few dots lie on or very close to the red line representing stable read current. Analogously, the read current at 1 s and 1.5 s (region II) is compared in figure 4.3, d). Despite the same interval of 500 ms, the cells are more stable compared to figure 4.3, c). The majority of dots is closer to the red diagonal. However, several cells still comprise current changes of up to 4 µA. Thus, the read variability decreases from region I to region II but is still present.
Nevertheless, it can be stated that the read current is characterized by highly stable log-normal distributions in region II after the initial relaxation in region I. Apparently contrasting to that, single cells show significant current fluctuations. The fact that the log-normal distribution heals after shaping suggests that the single cell fluctuations are not contrary to the stable distribution but in fact the origin of the intrinsic statistics. It seems that the apparently random fluctuations of single cells culminate in a dynamic balance represented by the log-normal statistics. Thus, removing "fail-bits" above a certain threshold current has a limited impact on the distribution, provided that the number of remaining cells is sufficiently large. Program-verify schemes may be beneficial to limit C2C variations, but the lowest achievable width of the read distribution is determined by read to read (R2R) variability due to the intrinsic statistics of the cell fluctuations. The physical origin of this phenomenon will be discussed in chapter 4.3.

HfO_2

Analogously, the read variability is studied for industrial HfO_2 based VCM devices. To determine the short term stability of the high resistive and the low resistive state, 10 k cells are programmed into LRS and HRS each. The initial read current after programming is given as dashed line in figure 4.4, a). It may be noted here that all experimental data obtained with these devices will be displayed with arbitrary units (a.u.) due to non-disclosure agreements. The LRS data displayed on the right side follows a straight line (regarding a linear current scale) and can thus be identified as a normal distribution. The HRS does not follow normal statistics but seems to be bended towards the read window. However, the HRS distribution becomes

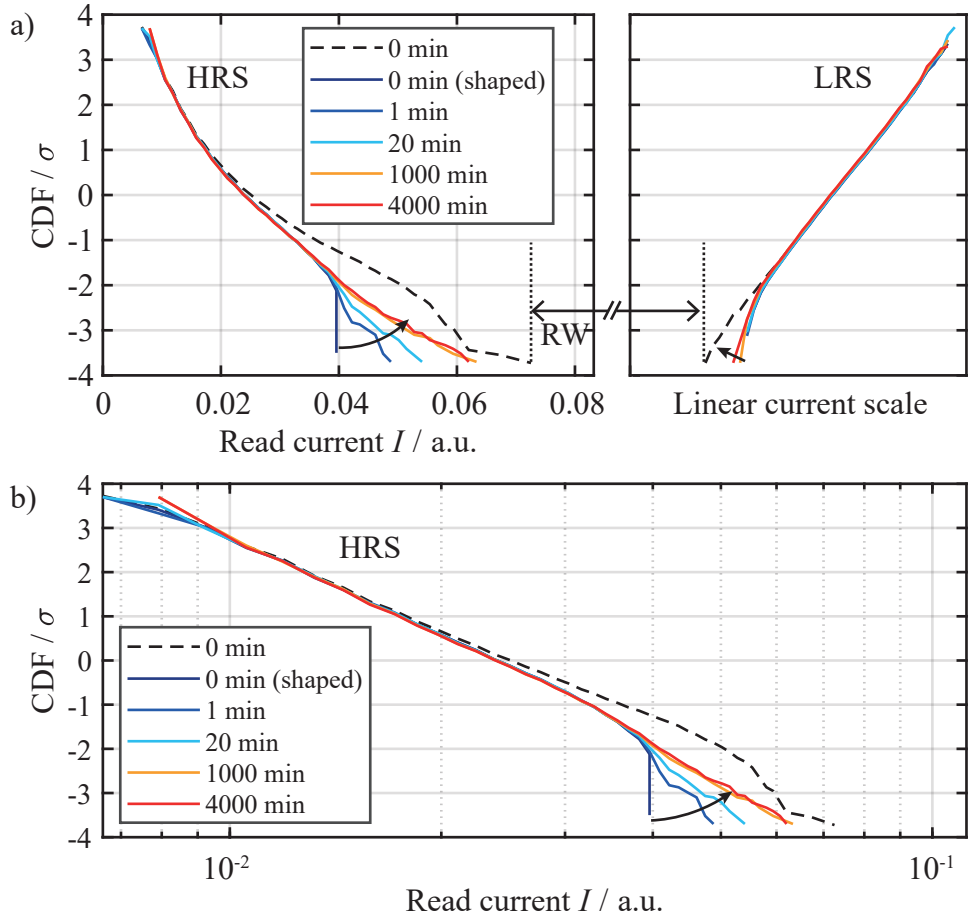

FIGURE 4.4: HRS and LRS read current distributions containing 10 k cells each. The HRS is inverted (complementary CDF) for better perceptibility of the read window (RW). Original data after programming displayed as dashed line. The same distribution after data shaping (HRS at 0.04 a.u.) is shown as solid blue line (0 min, shaped). a) linear current scale, b) logarithmic current scale identifying the volatile effect of shaping as reversion to the log-normal distribution.

linear if the current scale is set to logarithmic spaces, as depicted in figure 4.4, b). As mentioned above, this indicates that the high resistive state is log-normally distributed.

Equivalent to the experiments presented for the ZrO_2 devices, the effect of data shaping is investigated. Therefore, at $t = 0$ all HRS cells with a read current $I_{\text{Read}} > 0.04$ a.u. are discarded. As expected, the effect of this is volatile. On the linear scale it seems that a tail of "bad" cells developed already after 1 min delay which continues to grow until approx. 1000 min. With the logarithmic scale in figure 4.4, b) it becomes clear that this tailing is in fact a recovery of the log-normal distribution due to the intrinsic statistics.

The same data shaping experiment is performed with the LRS data by discarding all cells with I_{Read} below an LRS threshold. Here, a recovery of the intrinsic normal distribution is observed as well. However, the impact in the LRS is much lower

compared to the HRS. This is congruent with the common observation of less read noise in low resistive states compared to high resistive states [106, 116].
To study the origin of the intrinsic statistics, another 10 k cells are programmed into HRS by a single reset pulse (without shaping algorithm). Subsequently all cells are read once per second for a total duration of 20 ks. One exemplary device to device (D2D) distribution is given as dashed line in figure 4.5, a). As observed before, the HRS state is log-normally distributed. To clarify if the log-normal statistics can also be attributed to single cells, 20 exemplary read to read (R2R) distributions are added to figure 4.5, a). Here, the dashed line represents the read current of all 10 k cells at one distinct instant of time, whereas the other lines each represent the read current of one cell over all tested 20 k seconds.
It can be observed that the R2R statistics of individual devices are significantly different from the D2D statistics over all cells. The individual cells are characterized by a typically very steep slope, combined with few distinct current jumps. This indicates that two different classes of fluctuations occur in the R2R distributions, causing the low variability within the steep slopes and the higher variability given by the partially large current jumps.
Figure 4.5, b) shows the current traces of the six colored R2R distributions in figure 4.5, a). Equivalently to the R2R distributions, the read current traces comprise two different types of fluctuations. Firstly, all shown cells show current fluctuations with a high frequency but low amplitude around a very stable median. A R2R distribution over a section with only these jumps should be comparatively steep. But, secondly, few larger jumps seem to occur randomly, shifting the median with its smaller fluctuations to a different level. This means that the cell fluctuations within a certain current band account for the steep R2R distribution until the cell randomly jumps to a different current state.
Since the superposition of the read currents of all cells leads to a stable log-normal distribution, two different explanations come into account. On the one hand, the variability of the programming operation (e.g. reset) could cause the log-normal statistics (LNS), as the cells end up in random states which possibly follow a log-normal distribution. On the other hand, the random current fluctuations observed in the single current traces could generate a dynamic equilibrium following log-normal statistics. As already expected from the discussed shaping experiments, the first option seems unlikely, because it cannot explain the reversion of the shaped distribution into its log-normal shape.
This reversion is studied in more detail in figure 4.5, c). The black line represents the D2D distribution shaped at the first read by removing all cells exceeding 0.06 a.u. As expected, the remaining cells relax to the log-normal shape as shown by the gray

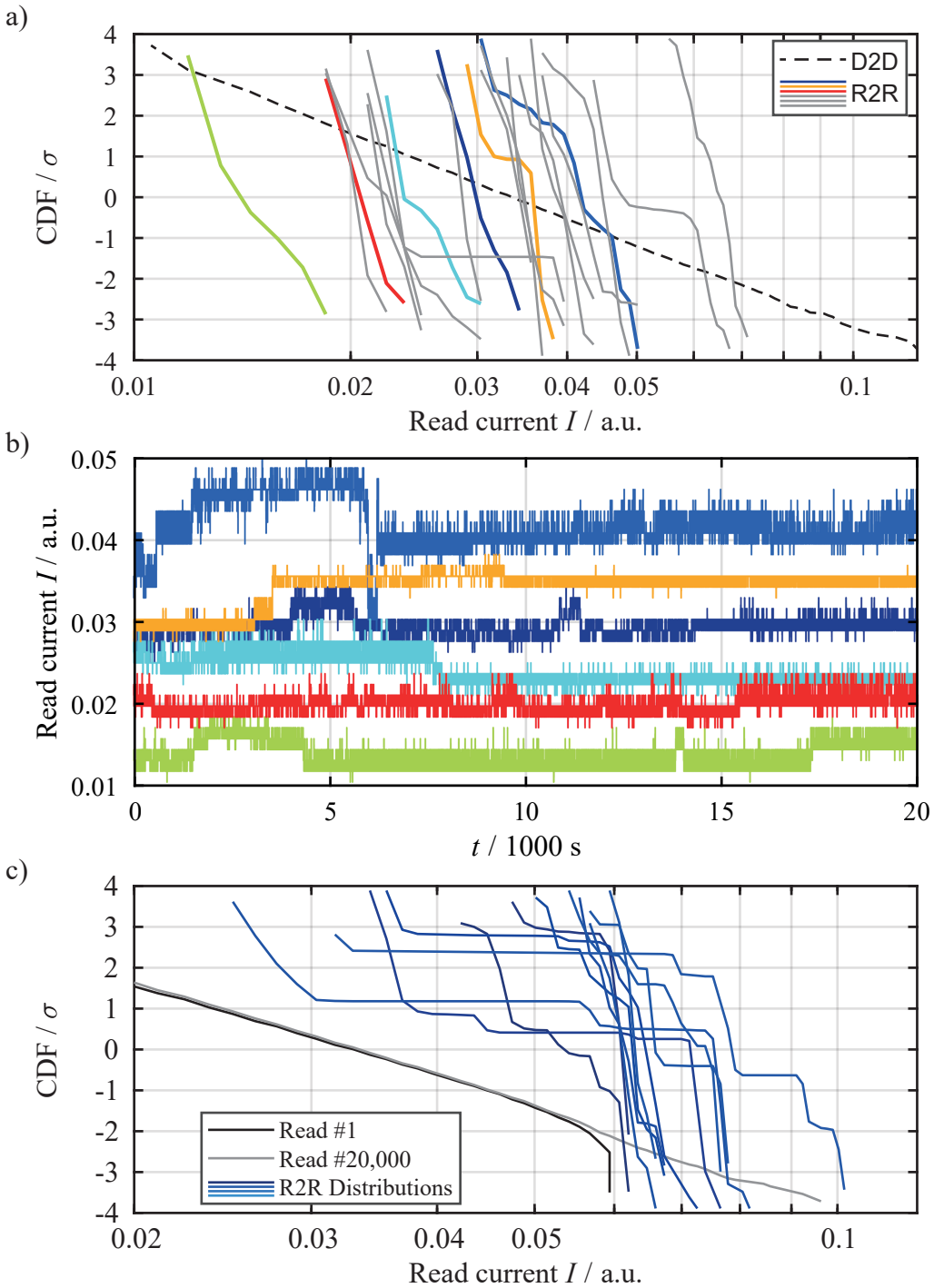

FIGURE 4.5: a) Device to Device (D2D) distribution compared to exemplary read to read (R2R) distributions. Whereas the distribution over all cells follows LNS, the individual cells are characterized by steep slopes and few random current jumps. The current traces for the 6 colored R2R distributions are shown in b). The read current typically fluctuates in a narrow band until random jumps occur. c) R2R distributions for cells which cause the reversion of a shaped distribution. The cut tail of the distribution is reverted by random jumps of cells from a broad range of the D2D distribution.

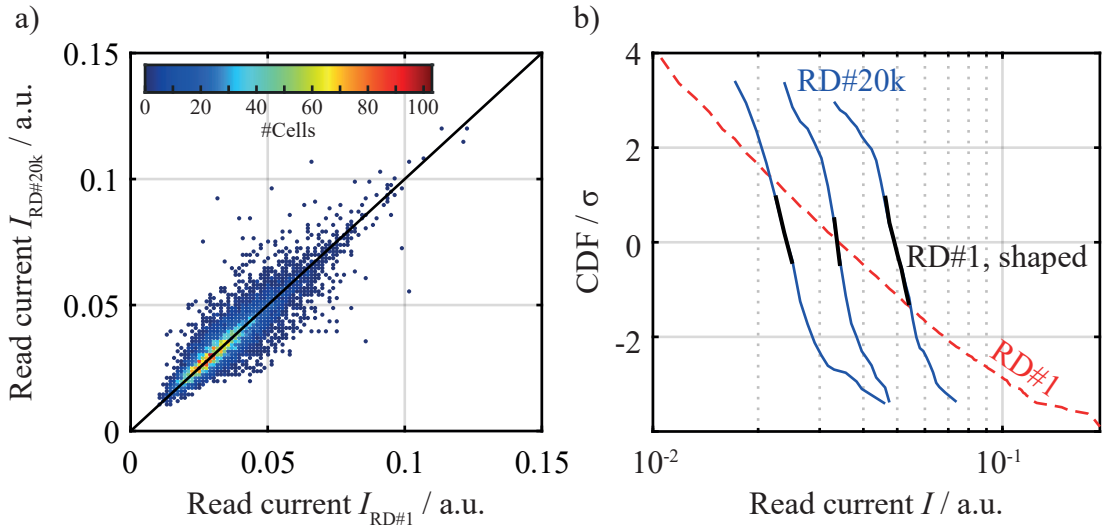

FIGURE 4.6: a) Lag plot comparing the read current of 10 k cells at the first read to the last read after 20 ks. The color marks the number of cells at the respective data point. Most cells lie close to the black diagonal representing no change between the two reads. However, current changes of up to 0.06 a.u. (> 100 %) are observed. b) Narrow sub-populations extracted for three current intervals from the first read. After 20 ks each group relaxed into a significantly broader distribution.

distribution for the last read after 20 ks. Here, several R2R distributions are shown in blue, each representing a cell which was below 0.06 a.u. at the first read and exceeded this threshold in the last read. It can be seen that the cells which are responsible for the reversion of the tail undergo partially very large current jumps.

This leads to several conclusions: It can be stated that the cells in the regrown tail come from a broad range of the original distribution. Additionally, the observed random jumps are responsible for the reversion of the distribution. Considering that these specific cells which jump into the cut-off section cannot "know" that the distribution was shaped before, it must be assumed that some of the cells with a read current above 0.06 a.u. at the first read would jump to a lower current (if not removed), because the over all distribution is observed to be stable. It seems that in each instant of time, approximately the same amount of cells jumps to a higher current as vice versa. These jumps seem to result in a dynamic equilibrium characterized by log-normal statistics.

For a better understanding of the occurring current fluctuations, a lag plot is given in figure 4.6, a) comparing the read current of all cells at the first read to their current at the second read. Although the log-normal distribution is observed to be stable, the individual devices change their state significantly. Current changes of up to 0.06 a.u. are possible which is more than 100 %.

Figure 4.6, b) shows another data-shaping experiment. Analogously to the study of Fantini et al. [86], the distribution of the first read (c.f. figure 4.5, c) is divided

into three narrow sub-populations. Within 20 ks the narrow sub-populations relax into the significantly broader distributions depicted by blue lines. Here, the shape is similar regardless of the current median of the extracted sub-populations. Each group seems to revert back towards the intrinsic log-normal statistics, enabled by the large current fluctuations shown in figure 4.6, a).
If the log-normal statistics were caused by the programming operation instead of by the observed current fluctuations, with the latter being an additional effect, it seems very unlikely that the distribution could be stable. Random current jumps of up to 100 % should disrupt the previously generated statistics. Therefore, it can again be stated that the random current fluctuations of individual cells are the most likely origin of the observed log-normal statistics.
It may be noted here that various methods exist to verify if a data set is normally or log-normally distributed [196]. Throughout this dissertation a graphical method is used. As explained in chapter 2.5, the normality of the distribution can be tested by plotting the normal quantiles of the data and comparing the resulting distribution to a straight line [197, 198]. This method provides a clear and simple to interpret way of recognizing intrinsic statistics [198]. However, additional methods could be used in order to verify the normality or log-normality of the data quantitatively. Commonly used examples are the Pearson Chi-Square test [199], the Shapiro–Wilk test [200] and the Jarque-Bera test [201]. Although the latter is frequently recommended for tests of normality, the performance of all statistical tests depends on the respective application and sample size [202, 203]. It is therefore suggested to try different approaches in the context of the given data. Moreover, extended or modified versions of the mentioned approaches might deliver the best performance [204]. As a consequence, it should be subject of future studies to identify the optimal test strategy as extension of the graphical approach for VCM ReRAM used in this dissertation.
In conclusion, the experimental investigation of read variability revealed intrinsic statistics of the programmed states, where the LRS is normally distributed and the HRS follows log-normal statistics. It is shown that these statistics are not primarily determined by the programming operations (although it might influence the read variability), but are an intrinsic phenomenon which can be attributed to random current jumps resulting in a stable distribution. In the following section, Factorial Hidden Markov Models are used to describe the read variability quantitatively. An explanation for the observed behavior will be given in section 4.3.

4.2.2 Factorial Hidden Markov Model

As proposed by Puglisi et al. [171], Factorial Hidden Markov Models (FHMM) are a powerful technique for the quantitative analysis of read noise data. FHMM methods do not deliver a physical understanding of the observed noise, but allow to decompose a noisy signal into multiple two-level noise signals described as individual Markov chains. Each chain is characterized by a jump amplitude and a set of transition probabilities. This enables a quantitative analysis of each occurring current change and gives rise to the distribution of jump heights and probabilities.

ZrO_2

Using the Baum-Welch algorithm introduced in chapter 2.3, the most likely set of four Markov chains is determined for all 540 HRS states recorded for the ZrO_2 based devices presented in section 4.2.1. Figures 4.7, a) and b) show results of the algorithm for the traces 2 and 5 from figure 4.2. The gray lines mark the recognized current levels which can be reached by superposition of the estimated Markov chains. It can be seen that the resulting current levels are in good agreement with the experimental data. Here, trace 2 in figure 4.7, a) is characterized by high frequent jumps of low amplitude whereas trace 5 in figure 4.7, b) comprises few large current jumps. Despite large differences in the general noise characteristics, the algorithm is observed to match the individual traces.
The evaluation shows that a maximum of four Markov chains is sufficient to generate the observed noise characteristics. A higher number of considered chains would not increase the precision of the evaluation since the additional chains would be characterized by very low amplitudes and therefore discarded. Furthermore, the computational burden increases drastically with increasing number of Markov chains. Although the FHMM evaluation does not provide a physical explanation of the analyzed noise characteristics, this observation indicates that a total number of 4 individual jump processes is sufficient to account for all observed current jumps in the investigated states.
Additional to the total number of jump processes, the FHMM analysis provides information about the occurring current jump amplitudes and their corresponding jump probabilities. With the respective sampling rate of the current trace, the latter can be translated into a jump frequency. Since the determined jump probability can be different for the two jump directions, the jump frequency is given by the rate limiting direction with lower jump probability. The resulting frequencies for the whole data set of 540 HRS states are shown as cumulative distributions in figure 4.7, c). Here, the data is divided into four groups with different jump amplitudes, revealing a trend

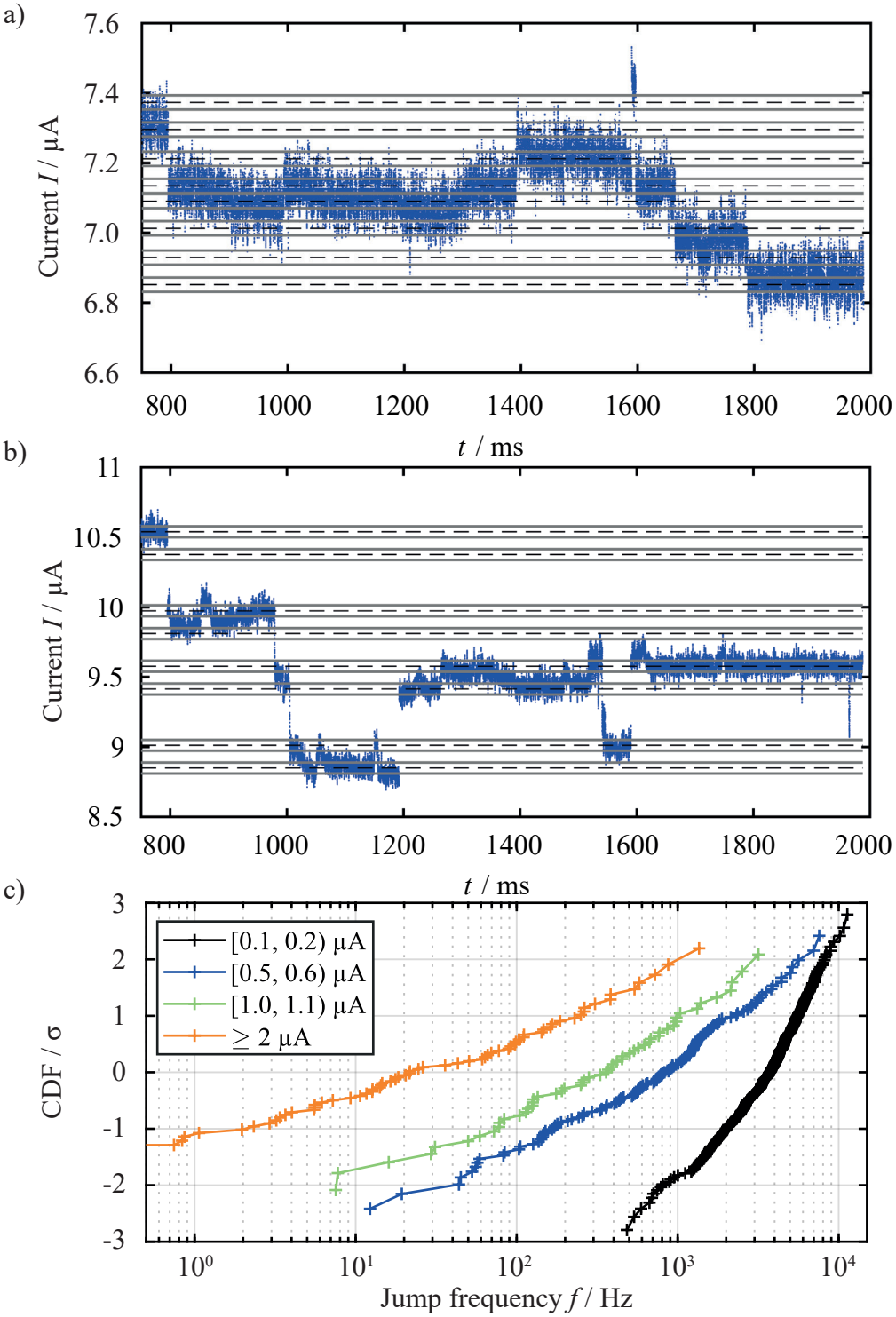

FIGURE 4.7: Exemplary fits obtained by FHMM methods for the current traces a) 2 and b) 5 from figure 4.2. The most likely Markov model is estimated considering four Markov chains. Gray lines mark the current levels reached by superposition of the respective Markov chains. c) Distributions of jump frequencies for different jump heights. Larger jump heights are observed to accompany lower frequencies. Reproduced with permission from S. Wiefels et al. [194], © 2020 IEEE.

of decreasing jump frequency with increasing amplitude. Whereas a median of 20 large jumps with $\Delta I \geq 2$ µA occurs per second, small jumps with $\Delta I < 0.1$ µA are two orders of magnitude more frequent.
The presented evaluation is used in [194] to develop a physics based kinetic Monte Carlo model explaining the observed current jumps. Here, it is demonstrated that the four groups of jumps with different amplitude can be assigned to distinct jumps of oxygen vacancies in the switching oxide. Lateral jumps of oxygen vacancies (perpendicular to the filament) only have a very small impact on the read current [194]. These $V_O^{\cdot\cdot}$ jumps explain the small current jump amplitudes given by the black line in figure 4.7, c). The higher current jump amplitudes are attributed to vertical $V_O^{\cdot\cdot}$ jumps (in filament direction). The KMC model revealed that the location of the $V_O^{\cdot\cdot}$ performing a vertical jump strongly influences the current change. It was found that jumps of $V_O^{\cdot\cdot}$ across the interface of plug and disc region have the highest impact on the measured read current [194]. In the same work, these findings are used to extend the JART VCM v1b compact model [205, 206] by including random jumps of $V_O^{\cdot\cdot}$ between plug and disc. It is shown that this jump type alone is sufficient to model the experimental read current distributions, including the reversion to log-normal statistics after shaping. This enables to model read noise with low computational burden for large scale ReRAM arrays which is of high importance for design strategies and reliability assessment [194].

HfO_2

Analogously to the results obtained for ZrO_2, the read noise data for the HfO_2 based devices given in figure 4.5, are evaluated using FHMM methods. The resulting distributions of the jump frequency for different jump heights is shown in figure 4.8. Although the current is given in arbitrary units, the same trend is observed. With increasing jump amplitude, the probability or frequency decreases. It may be noted that the sampling rate of this experiment is 1 S/s. Thus, the highest frequency to be resolved is 1 Hz and the extracted frequencies for different jump heights are on a significantly higher level than observed for ZrO_2. To determine the real frequency of the occurring current fluctuations, higher sampling rates would be beneficial.

4.3 Interpretation

Regarding the results presented above, it seems likely that the random current jumps are the origin of the log-normal distribution and responsible for the observed reversion of the distribution to its intrinsic statistics after shaping. To prove this

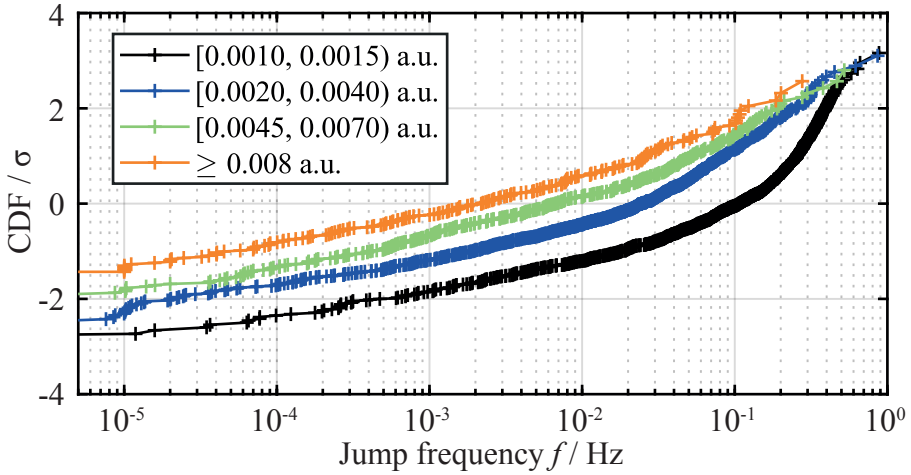

FIGURE 4.8: Results of the FHMM analysis for the HfO_2 based devices. Plotted are distributions of jump frequencies for different jump heights, evaluated for the read noise data from figure 4.5. Larger jump heights are observed to accompany lower frequencies.

assumption a simple model is set up: At first, 5000 virtual cells are initiated with identical read current of 0.035 a.u. which is approximately equal to the shaped distribution in figure 4.6, b) (middle). To each cell a random, normally distributed ($\sigma = 0.2, \mu = 0$) jump probability is assigned. Here, the absolute value of the random parameter is used to avoid negative jump probabilities. Within one standard deviation, i.e. 68 % of the cells, their jump probability is within the interval $[0, 20\ \%]$. Subsequently, 200 test cycles are performed where each cell may change its read current based on the assigned jump probability by a random factor. The latter is also drawn from the normal distribution with $\sigma = 0.05$. This means that 68 % of the jumps are in the interval $[-5\ \%, +5\ \%]$ of the read current before jump. The effect on the empirically generated read current distribution is demonstrated in figure 4.9. Starting from a vertical line at the initial read current of 0.035 a.u., the empirical distribution broadens with each cycle. The relaxed distribution from figure 4.6, b) is shown as dashed line and matched well by the simple model after approx. 50 cycles. After additional 150 cycles, the distribution approaches the log normal distribution in dotted lines which represents the original experimental distribution over all cells. Thus, the model proves that random jumps can be the origin of the observed log-normal statistics. In the following, it will be discussed where these jumps could originate and which impact the conduction mechanism might have.

Despite comprehensive research, the conduction mechanism of filamentary VCM ReRAM is still under debate. Considering ionic effects as origin of read noise, e.g. random jumps of oxygen vacancies in the conductive filament, the electrical conduction mechanism should play a major role in their effect on the read current. This also implies that the investigation of read noise might provide another clue of the

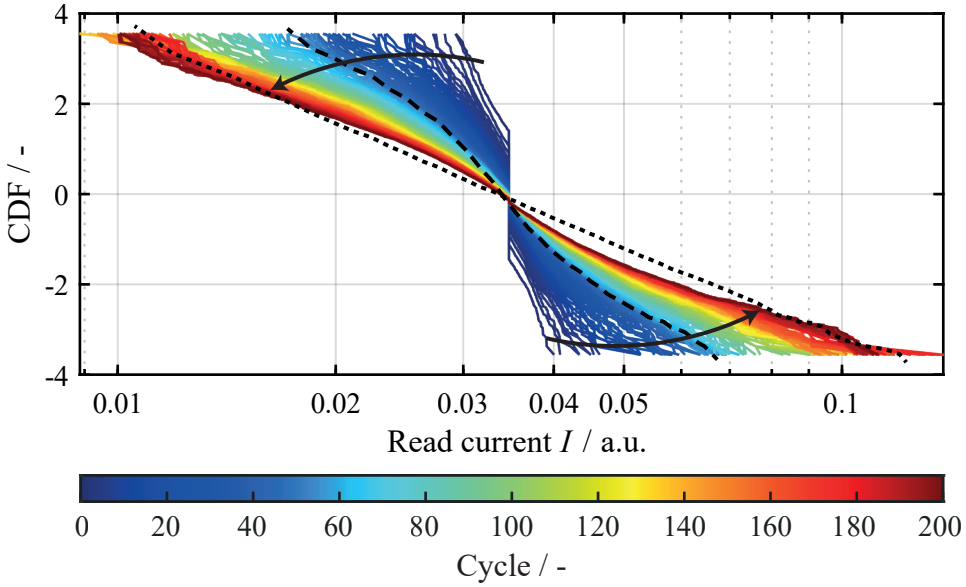

FIGURE 4.9: Empirical simulation of read current jumps. A theoretical distribution of 5000 cells with identical read current of 0.035 a.u. reverts towards the measured log-normal distribution via random current jumps. Thus, random jumps are likely to be the origin of the observed log-normal statistics.

effective conduction mechanism. As discussed in chapter 4.2.1, LRS and HRS are observed to have different noise characteristics. Firstly, read noise is significantly more pronounced in the HRS than in the LRS. Secondly, the intrinsic statistics of the distributions differ. Whereas the LRS follows a normal distribution, the HRS is characterized by log-normal statistics. As discussed above, these statistics are likely to result from the occurring random current jumps. If these current jumps originate from jumps of oxygen vacancies, their different impact on the read current suggests different conduction mechanisms for LRS and HRS.

Figure 4.10, a) shows exemplary $I - V$ characteristics of the investigated HfO_2 based devices. The sweeps are performed on one electroformed cell using a Keithley 2434B source measure unit with a sweep rate of 0.6 V/s and no applied current compliance. Instead a series resistance $R_S = 3\,k\Omega$ is used to limit the cell current during the set process. Figure 4.10, b) shows the same data plotted versus the actual cell voltage calculated by subtraction of R_S. According to the universal switching characteristics reported by Wouters et al. [207], the maximum SET and RESET current should be approximately equal. The significantly higher SET current in this experiment contradicts this expectation. A possible explanation might be leakage current along sneak paths during the SET process.

Nevertheless, from these $I - V$ characteristics the linearity of HRS and LRS can be determined. Whereas the LRS follows a straight line characterizing it as ohmic conductance, the HRS is observed to be non-linear and bended upwards. In figure 4.10, c)

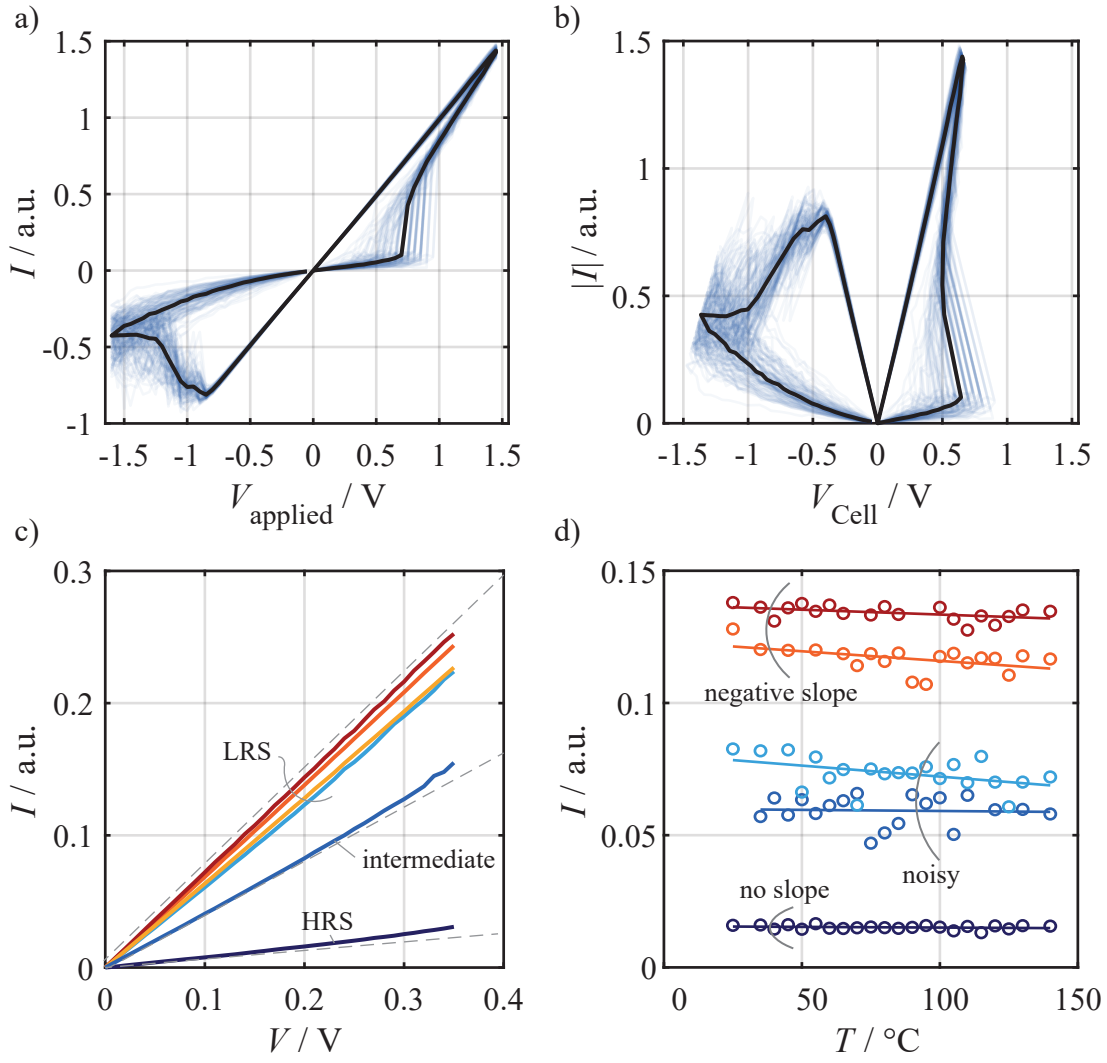

FIGURE 4.10: a) 100 $I - V$ sweeps performed on HfO_2 based cells with a series resistance of 3 kΩ close to the cell. The black line represents the current median at each voltage. b) Same data plotted versus the voltage V_{cell} dropping across the cell by subtracting the series resistance. c) $I - V$ sweeps with stop voltage below the switching threshold for different resistance states. The LRS is linear whereas the HRS and an intermediate state are bended upwards. d) Temperature dependence of the read current at 0.2 V for different resistance states. The LRS shows a negative slope with increasing T, whereas the HRS is constant. Intermediate states are too noisy to determine the temperature impact.

additional $I - V$ characteristics extracted below the switching threshold are given for different states. Again, the LRS can be described as linear, ohmic whereas the HRS is bended upwards. An intermediate state shows the transition between both characteristics. In conclusion, the different $I - V$ characteristics support the theory that different conduction mechanisms are present in HRS and LRS.

TABLE 4.1: Characteristics of HRS and LRS and likely conduction mechanism in HfO_2 based devices

Characteristic	LRS	HRS
Statistics	Normal distribution	Log-normal distribution
$I - V$	Linear	Bended upwards
T dependence	Negative slope	Independent of T
Likely conduction mechanism	Metallic	Tunneling

Another important aspect to determine the conduction mechanism is the temperature dependence of the respective state. The sub-threshold sweeps shown in figure 4.10, c) are therefore repeated at different cell temperatures ranging from room temperature towards 145 °C. For different states, the read current at 0.2 V is plotted versus the temperature in figure 4.10, d). The measurement reveals a slightly negative slope for cells in the LRS which suggests rather metallic behavior. In contrast to this, the read current in HRS seems to be very stable in this temperature range which could be attributed to a tunneling mechanism. The read current of intermediate states is superimposed by significant read noise at elevated temperatures. Therefore, the temperature dependence cannot be determined in this experiment. As summarized in table 4.1, the LRS follows normal statistics, shows linear $I - V$ characteristics and a slightly negative temperature dependence. All in all, this suggests a band transport of electrons resulting in a rather metallic conduction. Thus, random jumps of oxygen vacancies only have a low impact on the read current resulting in a narrow distribution. In contrast to this, the HRS is observed to follow log-normal statistics, shows non-linear $I - V$ characteristics and no temperature dependence. Although the investigation of the electrical signals is insufficient to determine the exact conduction mechanism (or multiple occuring mechanisms), these observations suggest some kind of tunneling processes limiting the effective read current. The exponential dependence of a tunneling process on the respective barrier provides a reasonable explanation of the log-normal statistics [124]. Considering a normally distributed tunneling gap induced by the random reconfiguration of oxygen vacancies, the exponential dependence of the resulting current leads to a log-normal distribution.

TABLE 4.2: Parameters of the empirical variability model

Parameter		ZrO_2	HfO_2
I_0	Fitting parameter	$6.00{\cdot}10^{-3}$ µA	$6.83 \cdot 10^{-5}$ a.u.
m^*	Electron tunneling mass	$9.10938 \cdot 10^{-31}$ kg	
Φ	Tunneling barrier	0.75 V	
V	Read voltage	0.35 V	0.2 V
μ_d	Gap median	1 nm	1 nm
σ_d	Gap standard deviation	30 pm	44.5 pm

4.3.1 Empirical Model

To prove the validity of the assumption that a tunneling current across a randomly distributed gap is the origin of the log-normal statistics, an empirical model is set up. Therefore, a tunneling current is defined as

$$I_\mathrm{Emp} = I_0 \exp\left(-2d_\mathrm{gap}\hbar\sqrt{2m^*(e\Phi - eV)}\right), \tag{4.1}$$

with the reduced Planck´s constant $\hbar$ and the elementary charge e. d_gap is the respective tunneling gap which is drawn from the normal distribution N according to

$$d_\mathrm{gap} \sim N(\mu_\mathrm{d}, \sigma_\mathrm{d}). \tag{4.2}$$

All parameters used in equations (4.1) and (4.2) are collected in table 4.2. Despite reasonable values being assigned to the physical parameters, it may be noted that the model cannot be used to determine an actual tunneling gap length. It just represents an empirical formalism to emphasize the concept of the read current being exponentially depending on a normally distributed parameter. Alternatively to a variation of the gap length, a normal distribution of the tunneling barrier height could be considered.

Using the presented empirical model, the read current distributions for both the ZrO_2 and the HfO_2 based cells are simulated. Therefore, random gap lengths are drawn from the normal distribution according to the fit parameters μ_d and σ_d in table 4.2. The resulting gap distributions are shown in figure 4.11, a) and b) for ZrO_2 and HfO_2, respectively. The resulting current distributions are calculated using equation (4.1) and shown in figure 4.11, c) and d). Here, the simulated distributions are compared to experimental data from section 4.2.1. The experimental data in figure 4.11, a) corresponds to the relaxed distributions in figure 4.3, a) obtained for

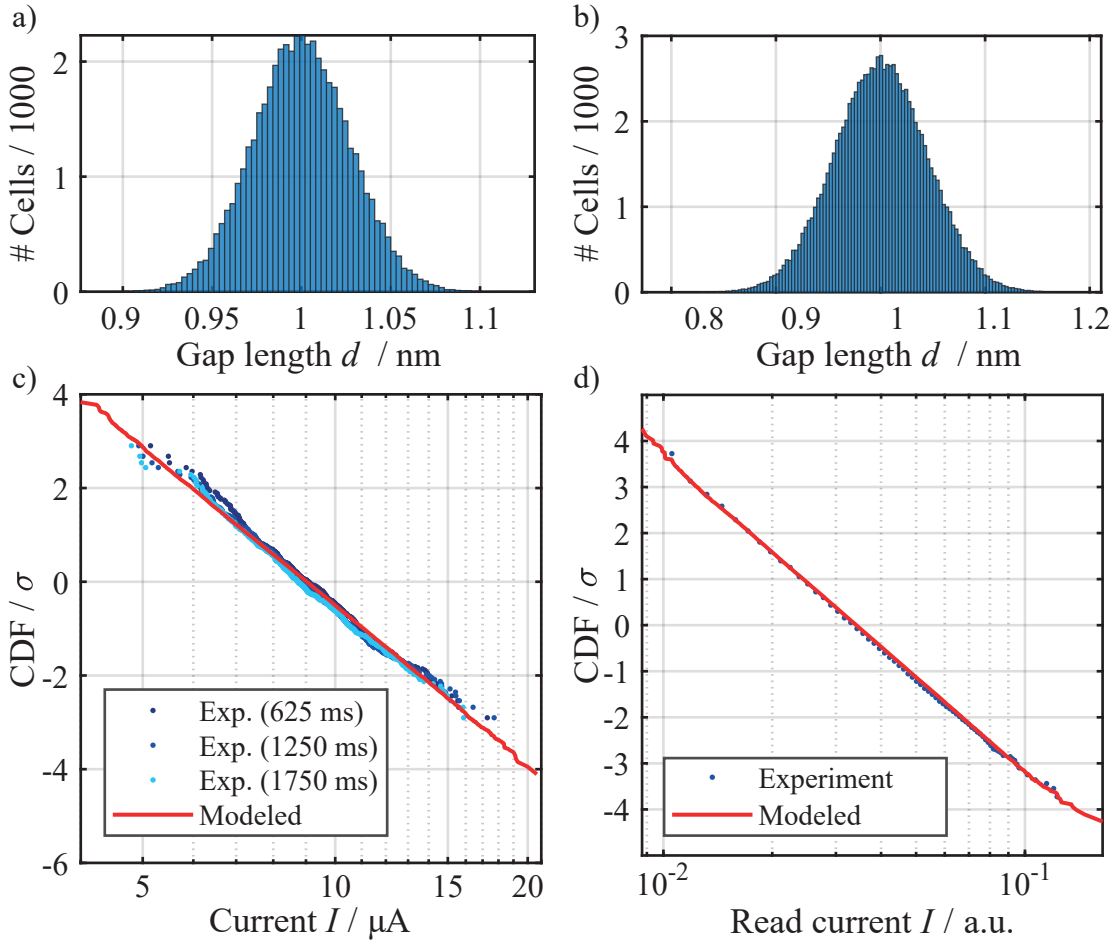

FIGURE 4.11: Results of the empirical variability model. a) Assumed distribution of the gap length d_{gap} according to equation (4.2) for ZrO_2. b) Respective distribution for HfO_2. c) Resulting, simulated current distribution for ZrO_2 in comparison to experimental read current distributions from figure 4.3. d) Simulated distribution for HfO_2 compared to experimental data from figure 4.5.

the ZrO_2 based cells. It can be seen that the simulated log-normal distribution is in excellent agreement with the experimental data.
Analogously, the modeled read current distribution in figure 4.11, d) is compared to the experimental R2R distribution shown in figure 4.5. Again, the empirical model is in excellent agreement with the experimental data.
In conclusion, the model demonstrates that the observed log-normal statistics can be explained by the dependency of the read current on a normally distributed parameter. Considering a normally distributed tunneling gap, the HRS current of both ZrO_2 and HfO_2 based devices are fitted. In chapter 5.2 this model will be extended to predict the retention characteristics of industrially relevant array scales. In the following section, statistical simulation methods will be used as an alternative approach to model the observed statistics.

4.3.2 Statistical Model

In the previous sections it was demonstrated that random jumps of oxygen vacancies are likely to cause the random current jumps observed in the read noise characteristics measured on typical VCM type ReRAM (c.f. figures 4.2, 4.5). To model such ionic reconfigurations, a spatial resolution of single oxygen vacancies is required which is not provided by the empirical model presented in the previous section or the JART VCM compact models [73, 205] mentioned above. In contrast, the KMC model developed by Abbaspour et al. [3–6], introduced in chapter 2.4 provides a sophisticated trap assisted tunneling solver for discrete defects within an oxide volume. This enables to investigate the impact of the reconfiguration of the defects which contribute to the electrical conductance of the cell.
Using the KMC model, read noise can be simulated by initiating a cell in the desired resistance state and subsequently simulating a constant voltage stress over a period of time at a certain temperature. Here, the KMC algorithm will determine and execute the most likely transition (generation, recombination or diffusion of defects, c.f. chapter 2.4) to occur in each increment of time, resulting in a change of the obtained read current. According simulations have been reported by Abbaspour in [5]. However, the KMC simulation induces a high computational burden which strongly limits the number of cells to simulate in a reasonable time.
To model read current distributions for industrially relevant array sizes comprising read noise, a more time efficient approach is required. Therefore, Kopperberg adapted the model of Abbaspour et al. [3–6], in order to account for the variability in large arrays of VCM ReRAM cells. This model discards the implemented KMC module and only uses the TAT current solver [208]. Furthermore, it is assumed that the spatial fluctuations of defects occur randomly in a confined volume. A sketch of the considered model is shown in figure 4.12, a). Within the 5 nm x 5 nm x 5 nm oxide, a filament volume is defined which is depicted by red edges. Between this filament box and the active electrode, a gap region is considered which contains no defects. As sketched by the simulation flow in figure 4.12, b) each simulated cell is initiated by placing a defined number of defects on random cites within the red filament box. Subsequently the Poisson equation (2.5) is solved for the generated configuration to obtain the respective potential which is then used to calculate the resulting read current using the TAT solver. After saving current and defect configuration, the simulation proceeds to generating the next cell by generating another random defect configuration [208]. Thus, the random reconfiguration of defects is modeled by random placement of defects within boundary conditions defined by the geometries of the filament volume. Since each distribution only has to be evaluated ones for their read current, a high number of random configurations can be evaluated

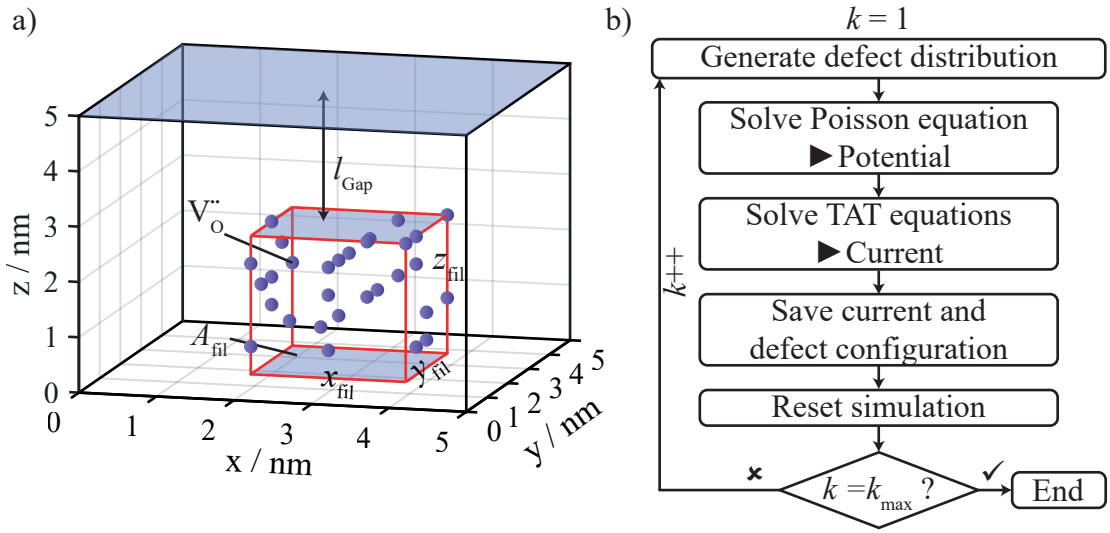

FIGURE 4.12: a) Schematic of the statistical variability model. Defects are placed on random cites within a defined filament volume with the dimensions x_{fil}, y_{fil} and z_{fil}. Between filament and active electrode, a gap region with the length l_{gap} is kept without defects. b) Simulation flow of the static model. Adapted from [208].

within reasonable time [208].

Using this model, read current distributions are calculated for theoretical VCM cells with different gap lengths l_{gap}, ranging from 0.75 nm to 3.75 nm. For each value of l_{gap}, 1000 cells are calculated and the resulting distributions plotted in figure 4.13, a). It is observed that the data for each gap length follows normal statistics as indicated by straight lines on a linear current axis. Since each of the simulated cells contains the exact same number of 20 oxygen vacancies, the different read currents within the individual distributions derive from different arrangements of the randomly placed defects. Starting at the largest gap length of $l_{gap} = 3.75$ nm, a decrease of this length results in a significant increase of the read current (despite constant number of defects). This increase can easily be understood by the lowered maximum tunneling gap between filament region and active electrode. Additionally, a decrease of the gap length decreases the slope and therefore increases the width of the distribution. This can be explained by the higher number of possible configurations in the enlarged filament volume. If the same number of oxygen vacancies is distributed over more available cites, more different combinations become possible and the difference between the best and the worst configuration (with respect to high read current) increases.

In contrast to the experimental data, the simulated distributions do not yet show log-normal statistics as it was expected for the HRS. The calculated read current varies with the random configurations of defects, but the model seems to provide no exponential dependence of the current on the defect configuration. However,

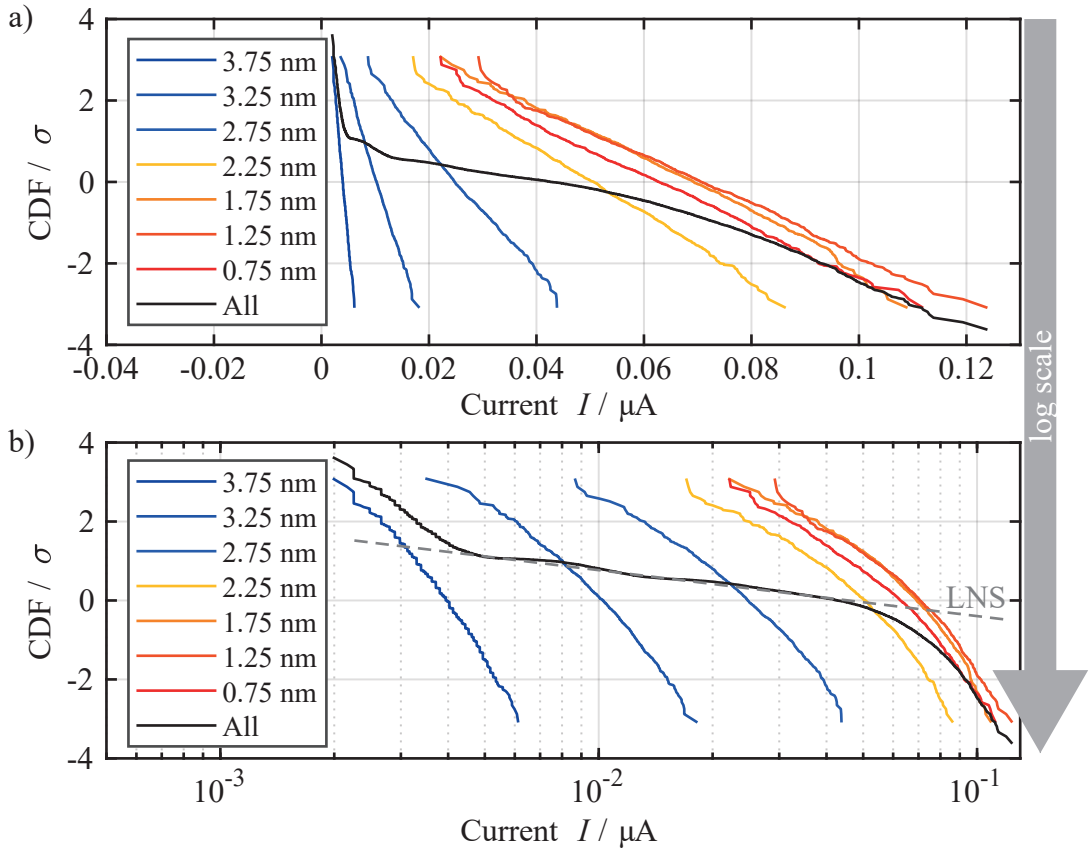

FIGURE 4.13: Simulated read current distributions using the statistical variability model. Calculated are read currents for randomly generated defect configurations for different lengths l_{gap} of the gap between filament and active electrode. For each value of l_{gap}, 1000 cells are simulated, resulting in a normally distributed read current. Log-normal statistics are observed if the data over all gap lenghts is combined (black line). a) Linear, b) logarithmic current scale. Adapted from [208].

as discussed before, a variability of the tunneling gap between filament and active electrode might result in LNS. Thus, the complete data set for varying gap lengths (0.75 nm - 3.75 nm) is combined to one distribution depicted by the black line in figure 4.13, a). By changing the current axis to a logarithmic scale, as shown in figure 4.13, b), it becomes clear that the combined distribution mostly follows LNS. At the borders defined by the maximum and minimum of the simulated gap lengths, the combined distribution deviates from the log-normal shape, due to the confined boundaries [208].

In conclusion, the statistical model shows that random fluctuations of the locations of oxygen vacancies may very well explain the read variability observed in read noise experiments. But, without a variation of the critical gap length, log-normal statistics can not be reproduced. This again suggests that a variability of the tunneling is the origin of the observed LNS in the HRS. Due to the exponential dependence of the read current on this gap length, the statistical model reproduces LNS if the simulation

takes a variability of this length into account.
The presented statistical model will be extended to model changes of the statistical parameters during retention experiments in chapter 5.3. Here, also the influence of the boundary conditions (filament geometry, number of oxygen vacancies) will be discussed in more detail.

4.4 Conclusion

This chapter demonstrates that read variability is one of the major challenges regarding the reliability of VCM ReRAM, as also reported in [86, 102]. Random current changes occur on various time scales. Considering large memory arrays, these current fluctuations determine the width of the effective read current (or resistance) distribution. Thus, read variability also limits the achievable read window. Since the fluctuations occur at any time, shaping algorithms are ineffective. Reprogramming or discarding cells which exceed a certain threshold, has no lasting effect because the distribution will relax towards its intrinsic statistics.
Using empirical models, random jumps of individual defects are identified as likely origin of the observed current fluctuations. Here, a resulting normal distribution of a read current limiting tunneling gap, seems to be the most probable explanation for the observed log-normal statistics in the HRS. In contrast, the conductance in LRS seems to be more metallic and thus the log-normal distribution does not occur. Statistical simulation methods support the conclusion of ionic movement being the origin of the read variability. Furthermore, the simulation reproduces LNS only if a variability of the tunneling gap is taken into account.

5 Retention

With VCM ReRAM being a promising candidate for future non-volatile memory applications, the long term stability of the programmed states is one of the key reliability aspects. Whereas several studies focus on the degradation of typical cells [36, 103, 130–133], it was already demonstrated that the median of a read distribution may degrade differently from a distribution tail [36]. This chapter will demonstrate that the most promising approach to understand and model retention characteristics is to focus on the whole statistics of the device under test.
The chapter is structured as follows: At first experimental retention results are presented for ZrO_2 based devices fabricated at *RWTH Aachen* and industrial HfO_2 based cells. Subsequently, the results are analyzed quantitatively and modeled by two different approaches: An empirical tunneling model is presented which describes the retention of the HfO_2 based cells and allows for extrapolation towards different temperatures and times. This model is published in [209]. Additionally, a statistical model is introduced which explains the observed effects and is published in [210]. Finally, the statistical model is extended to a dynamic version which is able to explain the coexistence of low short term stability (Read Noise) and high long term stability (Retention) which remains one of greatest open questions regarding VCM ReRAM.

5.1 Experimental Results

In the following, experimental retention results are presented for the two investigated material systems.

5.1.1 ZrO_2

At first, the long term stability of the standard ZrO_2 devices is studied. Since typical retention requirements of 10 years at 85-125 °C [82] cannot be tested directly, retention experiments are typically performed at elevated temperatures. Therefore, three groups of cells are programmed to HRS and three groups to LRS. The groups are programmed on three different dies of the same stack and manufacturing. Here, each die contains one HRS and one LRS group of at least 200 single cells. All cells are programmed by initial electroforming, followed by 10 switching cycles. All

operations are performed by a program-verify approach as explained in chapter 3.3, with a target read current $I_{\text{read}} < 10$ µA for RESET and $I_{\text{read}} > 60$ µA for SET. Here, all programming pulses are triangular with a rise and fall time of 2.5 ms each. The resulting sweep rate depends on the respective stop voltage which ranges

- from 2.0 V to 4.0 V during forming,
- from 0.6 V to 1.4 V during SET, and
- from -1.3 V to -2.0 V during RESET.

All voltages are applied to the Ta top-electrode and the Pt bottom-electrode is connected to ground. It may be noted here, that the typical voltages, where the switching operation is successful are 3.2 V for forming, 1 V for SET and -1.8 V for RESET.
After programming, all cells are read by a rectangular read pulse with a width of 20 µs at 0.2 V. The resulting read current distributions are depicted in figure 5.1 by dark blue circles (0min). As expected from chapter 4, the HRS read current follows log-normal statistics whereas the LRS read current is normally distributed. Caused by the program-verify algorithm, the LRS distributions show a bending or limitation at the SET programming target of 60 µA. Cells with a lower read current received additional SET pulses which shaped the distribution away from the intrinsic normal statistics. Regarding the HRS, few cells exceed the RESET target of $I_{\text{read}} < 10$ µA. As demonstrated in chapter 4, significant random current fluctuations in the HRS lead to a fast reversion to the intrinsic log-normal statistics.
After the initial read out (0min), all three dies are baked at 150 °C, 175 °C or 200 °C and read out after different bake times with largely logarithmic spacing. For each read the die is removed from the oven and contacted again by the probe card. Here, the time to cool down from baking temperature to room temperature may distort the total bake time. Although the effect may be negligible at long bake times, it has to be considered for short increments (1 min to 10 min). Furthermore, re-contacting the cells for each read-out may alter the measured state due to mechanical stress or electrical discharge. In all six groups, some cells were shorted from read to read which is likely to be caused by contacting issues.
All read distributions are fitted regarding their intrinsic log-normal (HRS) or normal (LRS) statistics. The fitted distributions are depicted by the solid lines in figure 5.1. Despite the mentioned issues, a general trend is observed with all dies for both HRS and LRS. In each case, the median of the distribution shifts towards lower current with increasing bake time. Whereas this may cause a retention failure of the LRS, a shift to lower current is beneficial for the HRS. However, additional to the shifting, all distributions tilt towards higher standard deviation, i.e. broadening of

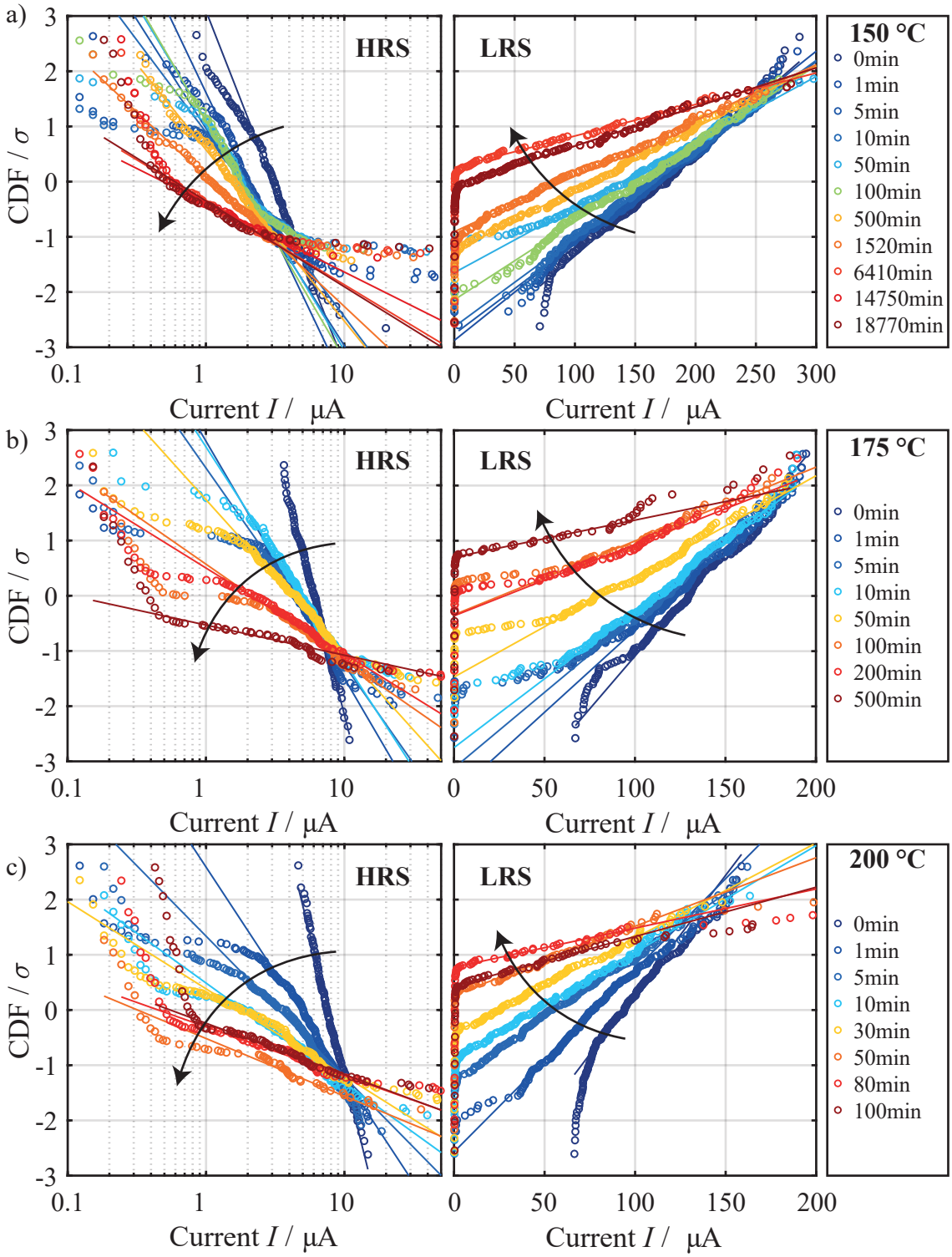

FIGURE 5.1: Retention results for Pt / ZrO_2 / Ta / Pt devices. On three dies one group of cells is programmed to HRS and one to LRS. Each group contains at least 200 cells. All cells are read initially (0min) and after logarithmically increasing time steps at three different temperatures. a) 150 °C for up to 13 days. b) 175 °C for up to 8 hours. c) 200 °C for up to 100 minutes. All distributions are fitted for their intrinsic log-normal (HRS) or normal (LRS) statistics. Data depicted by circles, fits by solid lines. In each case and for both HRS and LRS the distribution shifts towards lower current and tilts/broadens towards higher standard deviation.

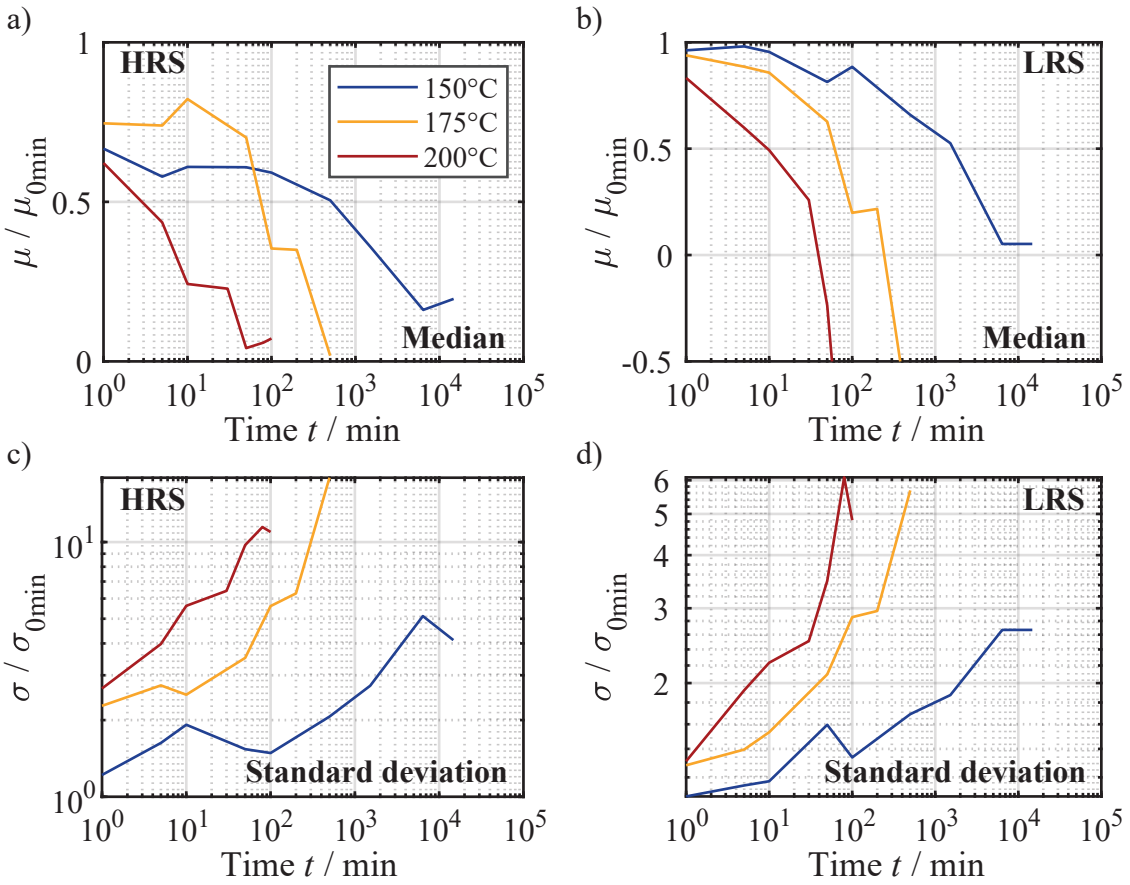

FIGURE 5.2: Characteristic parameters of the log-normal (HRS) or normal (LRS) distributions fitted in figure 5.1, normalized regarding the initial read (0min). a) Degradation of median μ in HRS. Below 1 h of bake time, the trend is disturbed by experimental artifacts. Later, the degradation is accelerated by increased temperature. b) Degradation of median in LRS. Negative values of μ occur if more than 50 % of cells fall below the current resolution limit. A negative μ cannot be interpreted physically as negative current median but describes the statistics of the fitted cells at higher percentiles (c.f. figure 5.1). Significant degradation of the standard deviation σ in HRS c) and LRS d) is identified as main origin of read window closure.

the distribution. Despite decreasing median current in the HRS, the cells at lower percentiles (often *tail-bits*) decrease the effective read window between HRS and LRS due to broadening of the whole distribution. Regarding the LRS, both shift and tilt are harmful for the read window.

To evaluate these trends quantitatively, the characteristic parameters of the log-normal (HRS) or normal (LRS) distribution fits are extracted and plotted in figure 5.2. Here, all parameters are normalized for the initial read (0min). As stated above, a general trend of decreasing current median μ and increasing standard deviation σ is observed for both HRS and LRS at all tested temperatures. With increasing temperature the degradation of both μ and σ is accelerated as expected for a temperature activated process. Only within the first hour of bake time, the trend of the HRS

current median (figure 5.2, a)) is disturbed. The reason for this might either be the mentioned contacting issues or the inaccuracy anticipated for shorter bake times.
With degradation of the LRS, the shift and tilt of the distributions leads to an increasing number of cells with very high resistance and thus read currents below the resolution limit of the read operation. Since these cells cannot continue the normal statistics of the main distribution but seem to accumulate close to 0 µA, the current median of the fitted distributions eventually becomes negative, as soon as more than 50 % of cells are accumulated close to 0 µA. This is observed in figure 5.2, b) for 175 °C and 200 °C. Although negative values of μ cannot be physically interpreted as median of the read current, the parameter still describes the statistics of the sub-population of cells above the resolution limit. Thus, the percentage of cells which failed a defined retention criterion may very well be drawn from this fit. Furthermore, the fit of the higher percentiles still allows to quantify the degradation of the initial read current distribution. Despite the physically nonsensical median μ, the standard deviation σ extracted from this fit provides a measure of the width of the distribution.
In conclusion, the evaluation of the retention by the trend of single cells is ineffective. The results demonstrate that the degradation of the programmed states directly affects the intrinsic statistics, i.e. shift of the current median μ and tilting / broadening with increasing standard deviation σ. In order to develop an extrapolation model for the retention at operating temperatures, the presented data suggests to focus on statistics instead of individual cells. This will be utilized in section 5.2. Furthermore, the broadening of the read current distribution is identified as major problem for a stable read window.

5.1.2 HfO_2

Additional to the presented data for ZrO_2, experimental retention data is provided by *Infineon Technologies*. The measurements are conducted on industrial ReRAM devices based on HfO_2 as switching oxide.

Measurement Procedure

Using a dedicated test setup, the devices are (after initial electroforming) cycled 1000 times. Subsequently, one half of the devices is programmed into LRS, the other half into HRS. All programming steps are performed at room temperature and a program-verify algorithm is used, i.e. after each pulse (forming, set or reset) the device is read and receives additional pulses if necessary. With each step, the pulse width and height are adjusted to ensure reliable switching with minimal stress to the device. Each group (LRS and HRS) contains 2.5 M cells. After programming

the devices are read out initially. To determine the stability of the hereby obtained read current distribution regarding time and temperature, the sample is exposed to an elevated temperature (bake) and read out again. This process is repeated at the same temperature for increasing periods of time. All reads are performed at room temperature. To extract the impact of time and temperature on the read current distributions, this procedure is repeated for four different bake temperatures, 150 °C, 175 °C, 215 °C and 260 °C, with an explicit set of 2.5 M devices each. Although the highest temperature (260 °C) is especially relevant regarding soldering robustness, the aim is to extract a trend at elevated temperatures and estimate the long term stability of the devices at the respective maximum operating temperature.

General Observations

Figure 5.3 exemplarily shows read current distributions for a retention experiment at 175 °C. As can be seen from figure 5.3, a), the resulting LRS distributions are observed to be very stable (not shown), whereas the HRS changes over time at elevated temperature. Therefore, this dissertation focuses on the latter and all further figures in this chapter will exclude LRS distributions.
The black line in figure 5.3, a) represents the initial HRS read current distribution. After the sample is exposed to 175 °C for 145 h, the distribution changes in its shape and seems to develop two sub-distributions. One sub-population shifts to lower and the other shifts to higher read current. In figure 5.3, b) the same data is plotted on a logarithmic current scale. Here, the initial distribution as well as the one after bake are composed of two parts: The linear part follows log-normal statistics as described in chapter 4. At higher read current the distribution deviates from this behavior and exhibits a current limiting effect. Although this bending is beneficial for the read window, it contradicts the observation that the HRS of HfO_2 (or ZrO_2) based ReRAM intrinsically follows log-normal statistics. An explanation for this phenomenon is given by the statistical model in chapter 5.3.
With the logarithmic current scale (figure 5.3, b)), the general shape of the distribution is not altered by baking. Instead of developing sub-populations, the log-normal distribution after bake is tilted with respect to the initial distribution. Since the slope of the log-normal distributions is one of its two characteristic parameters (cf. equation (2.28)), it seems reasonable to interpret the impact of the bake as a change of the statistical parameters instead of a shift of the current of individual devices or sub-populations.
To further emphasize this, the devices are read out twice after bake (145 h at 175 °C) within 6 h at room temperature. As can be seen from figure 5.3, a) and b), the distribution is stable between the first and second read and the lines coincide. Figure 5.3, c)

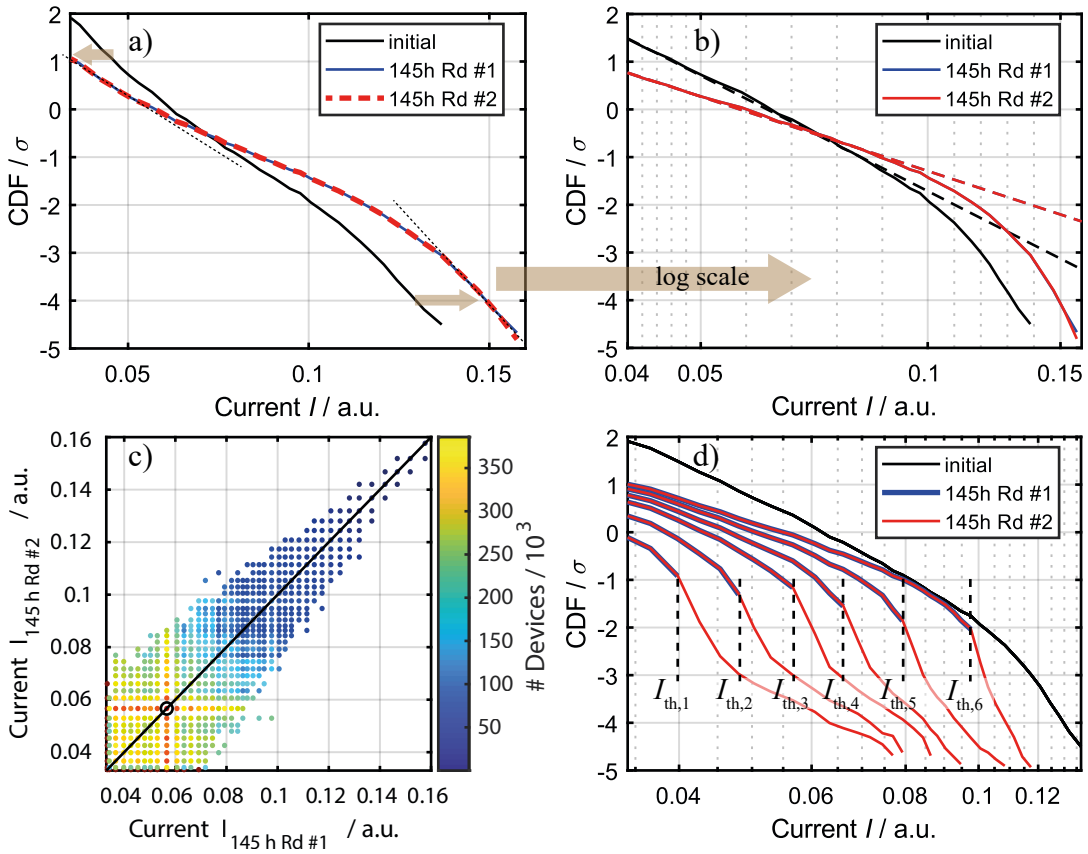

FIGURE 5.3: Experimental HRS Retention Data. a) Distribution of the initial read current and after 145 h at 175 °C (read twice within 6 h at room temperature). The linear current scale suggests a division into two sub-populations during bake. b) Log(I) scale reveals log-normal statistics with current limiting effect before and after bake. c) Time lag plot of first and second read after bake showing significant fluctuations of individual cells despite stable distribution. d) Data shaping experiment for different threshold currents $I_{\mathrm{th,i}}$: Cells exceeding the threshold in read 1 are removed from the data set. The tail re-evolves in read 2. Partially reproduced with permission from S. Wiefels et al. [209] and [210], © 2020 IEEE.

shows a lag plot comparing the current of the first and second read. Seemingly in contrast to the stable distribution, it can be seen that the individual cells change their state significantly within the two reads without bake in between. As discussed in chapter 4, it is assumed that the inherent current fluctuations of all individual cells induce the observed log-normal statistics. Additionally, a data shaping experiment is conducted by removing all cells from the data set which exceed a certain threshold current in the first read. The identical cells are then also removed from the data set of the second read. If all individual cells were stable, the shaped distributions of the first and second read should still coincide. Figure 5.3, d) shows the result of this experiment for different threshold currents $I_{\mathrm{th,i}}$. As expected from the results in chapter 4, individual cells are not stable and the removed tail of the distribution re-evolves until the second read. Again it has to be stated that shaping algorithms

have a limited effect and the intrinsic statistics have to be considered.
In conclusion, it can be stated that the evaluation of the retention experiment has to be based on the parameters of the distribution with its underlying log-normal statistics instead of individual cells. Therefore, the following evaluation of the experimental retention data will focus on the characteristic parameters μ and σ of the log-normal distribution (cf. equation (2.28)).

Quantitative Analysis

In order to perform a quantitative analysis of the degradation, retention experiments are performed at four different temperatures, ranging from 150 °C to 260 °C. In each experiment 2.5 M cells are programmed into the HRS. The read current is read initially and after several periods of baking time. Figure 5.4 shows the resulting read current distributions. Up to a certain threshold of approx. 0.1 (a.u.) all distributions are linear on the logarithmic current scale and therefore log-normally distributed. This part is fitted and evaluated quantitatively in the following. Since the general shape of the log-normal distribution stays intact, two effects are possible:

(i) Shifting of the median μ towards higher or lower current, or

(ii) tilting of the distribution, resulting in a change of the standard deviation σ.

The experiments at all four baking temperatures show qualitatively the same trend. The dominating effect of the bake at each temperature is a tilting (ii) of the distribution towards higher σ. This means that the width of the distribution increases over time. In each case, the tilt is most pronounced during the first increment of time ranging from 15 minutes at 260 °C to 1 hour at 175 °C. The tilt is superimposed by a shift of the median μ (i). Here, a shift to lower current is observed within the first bake. Subsequently the direction of the shift changes resulting in a monotonous increase of the read current. To gain a quantitative understanding of these trends, the characteristic parameters μ and σ are extracted from the linear fit in figure 5.4 and depicted in figure 5.5. As shown in figure 5.5, a) the standard deviation σ which represents the reciprocal slope of the distribution increases strongly in the beginning, followed by a monotonous increase with decaying derivative. figure 5.5, b) depicts the course of the median μ. As mentioned above, the current initially decreases. Subsequently μ increases monotonously. In order to resolve the trend during the first hours of bake, figure 5.5, a) and b) only show the first 15 hours.
For both parameters, the choice of an appropriate time scale reveals a clear trend which may give rise to the underlying physics of the respective process. As can be seen from figure 5.5, c) the standard deviation σ perfectly follows a logarithmic time

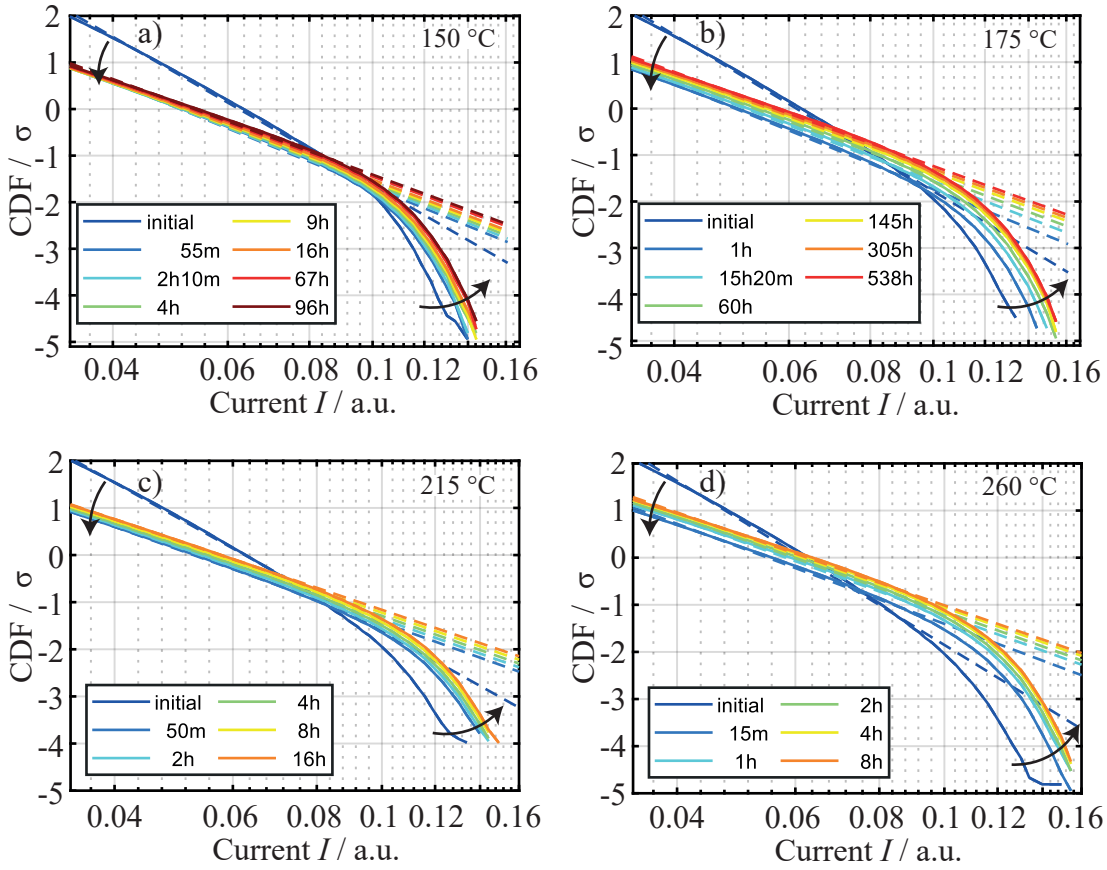

FIGURE 5.4: Read current distributions obtained by retention experiments at a) 150 °C, b) 175 °C, c) 215 °C and d) 260 °C. All distributions show a tilt and shift during bake, whereas the general shape stays intact. Partially reproduced with permission from S. Wiefels et al. [210], © 2020 IEEE.

dependence. As shown in figure 5.5, d) the median μ increases linearly with $t^{1/4}$ if the initial decrease is excluded. This decrease can not be resolved in the given data set due to the limited temporal resolution. Though, the linear slopes (versus $t^{1/4}$) of μ for 175 °C to 260 °C can be extrapolated to a common origin at 0.048 (a.u.), as depicted by the dashed lines in 5.5, d). This could represent the terminal point of the initial decrease. However, at 150 °C this extrapolation does not fit, the initial decrease of the read current is terminated earlier. This indicates that the underlying process is temperature activated. In combination, it seems that μ is affected by two competing processes. The process dominating initially (1) results in a current decrease, the other (2) in a current increase (linearly with $t^{1/4}$). At 150°C the transition point is resolved at approx. 3 hours. At higher temperatures the transition point is shifted into the region between the initial read and the read after the first bake. The fact that later only the second process is observed suggests that the first process is self limiting at a certain point. This interpretation is supported by the common origin of the extrapolation lines mentioned above. Due to the limited resolution in the region of

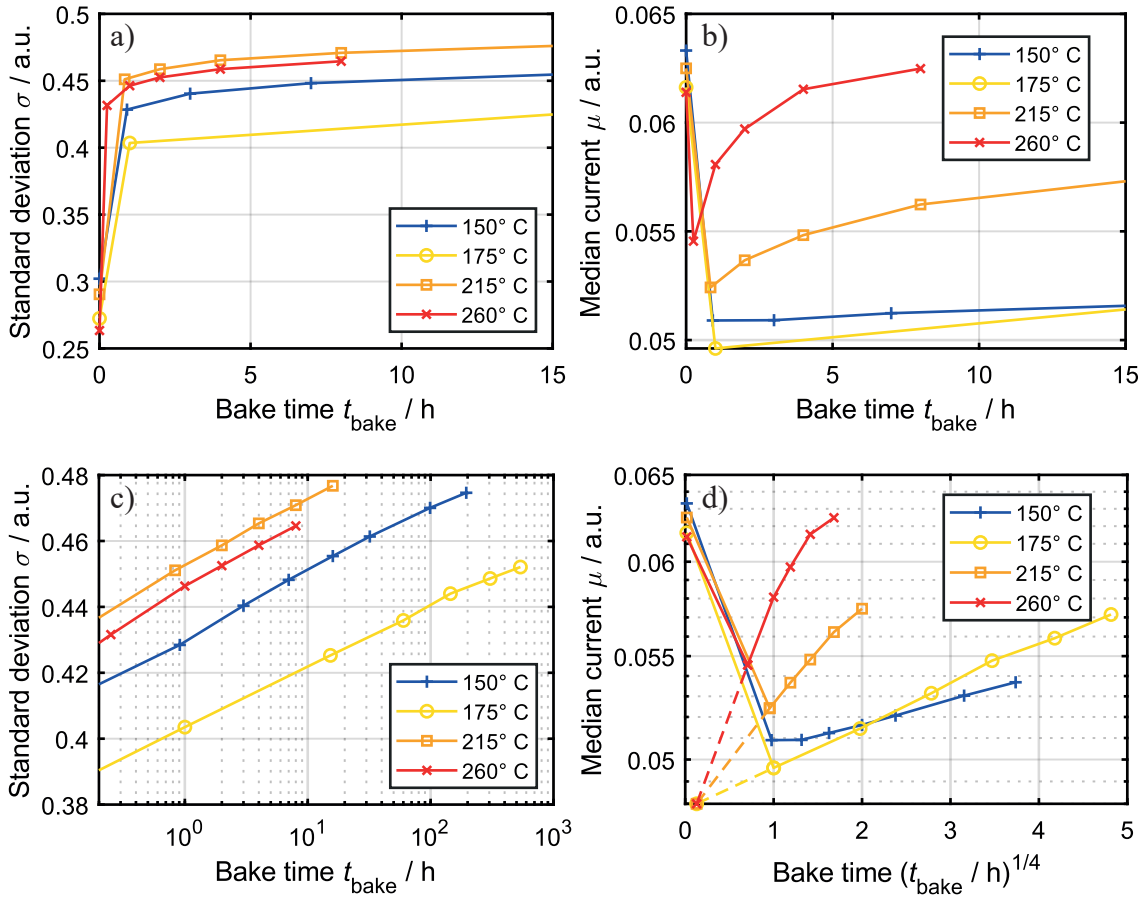

FIGURE 5.5: Characteristic parameters μ and σ of the log-normal distribution, extracted from linear fits in figure 5.4. a) and b)show σ and μ for the first 15 hours on a linear time scale; c) σ linearly increases with $\log(t_{bake})$, d) μ initially decreases, subsequently increases linearly with $\mu \propto t^{1/4}$. Partially reproduced with permission from S. Wiefels et al. [210], © 2020 IEEE.

process ① the quantitative analysis in this work focuses on process ②.
With respect to the application, the aim in optimizing the retention of a device is to ensure a sufficiently large read window between the lowest percentiles of the HRS and LRS distributions. Therefore, the trace of the current close to the read window during a retention experiment is of particular interest. Thus, the current at -4σ is extracted and depicted in figure 5.6 on two different time scales. Figure 5.6, a) shows the course of the first four hours of bake time on a linear time scale. The initial decrease of the current, as it is observed for the median μ, is not observed at this end of the distribution. Instead the HRS current close to the read window increases monotonously. Here, the steep increase suggests that the current at -4σ is predominantly affected by the tilt of the distribution during the first hours of bake time. Later, the superposition of tilt and shift of the distribution determines the current at -4σ as depicted in figure 5.6, b). This indicates that a quantitative evaluation of the current at -4σ is challenging. Instead, understanding the trend

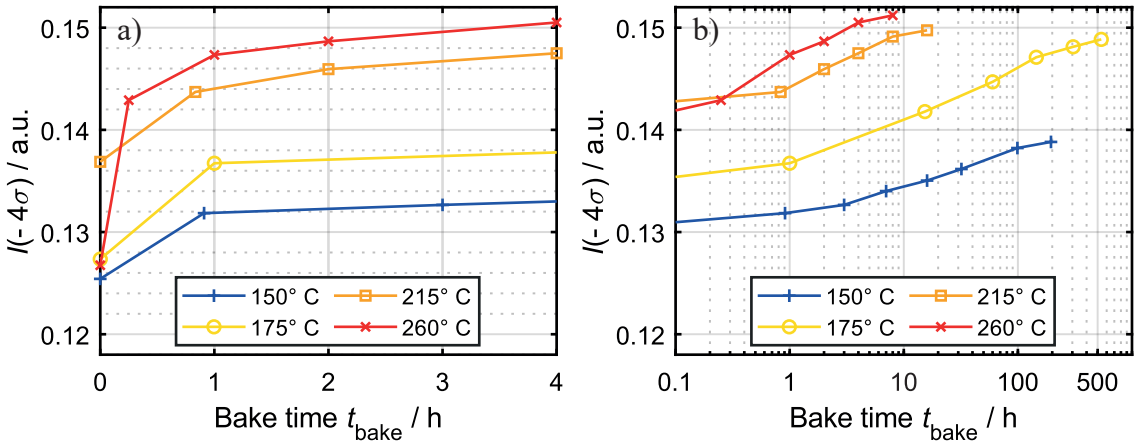

FIGURE 5.6: Read current close to the read window (-4σ). a) Linear plot of the first hours reveals current increase due to tilting of the distribution. No initial decrease is observed. b) The trend of $I(-4\sigma)$ is a superposition of the trend observed for μ and σ (cf. figure 5.5). Adapted from [210], © 2020 IEEE.

of the characteristic parameters μ and σ of the log-normal distribution could be the key in predicting the course of the current close to the read window. Though, this evaluation of the current at -4σ reveals that the tilting (or broadening) of the log-normal HRS distribution might be the major challenge in ensuring a stable read window.

Activation Energy

Another widely accepted approach to estimate the retention time at operating temperatures is to extract the activation energy of the thermally activated process which causes the degradation [82, 145–147]. Assuming that this activation energy is constant over the temperature range between operating and testing temperatures, the high temperature degradation can be extrapolated to the lower operating temperatures. The first challenge of this approach is to reliably determine an activation energy from the measured quantity. It is questionable whether the activation energy of a degradation mechanism like (e.g.) diffusion of defects is directly reflected by a read current or resistance. Furthermore, it is conceivable that multiple processes are involved in the degradation which cannot be distinguished in the electrical signal.
Figure 5.7 shows different approaches to extract the activation energy from read current distributions. Here, another data set is used which is obtained analogously to the data discussed before, but covers five bake temperatures, ranging from 125 °C to 250 °C. The read current distributions (not shown) are qualitatively identical to the preceding.
As outlined before, the critical region of the distribution is the current close to the read window. Thus, figure 5.7, a) depicts the degradation of the current at -4σ in percent.

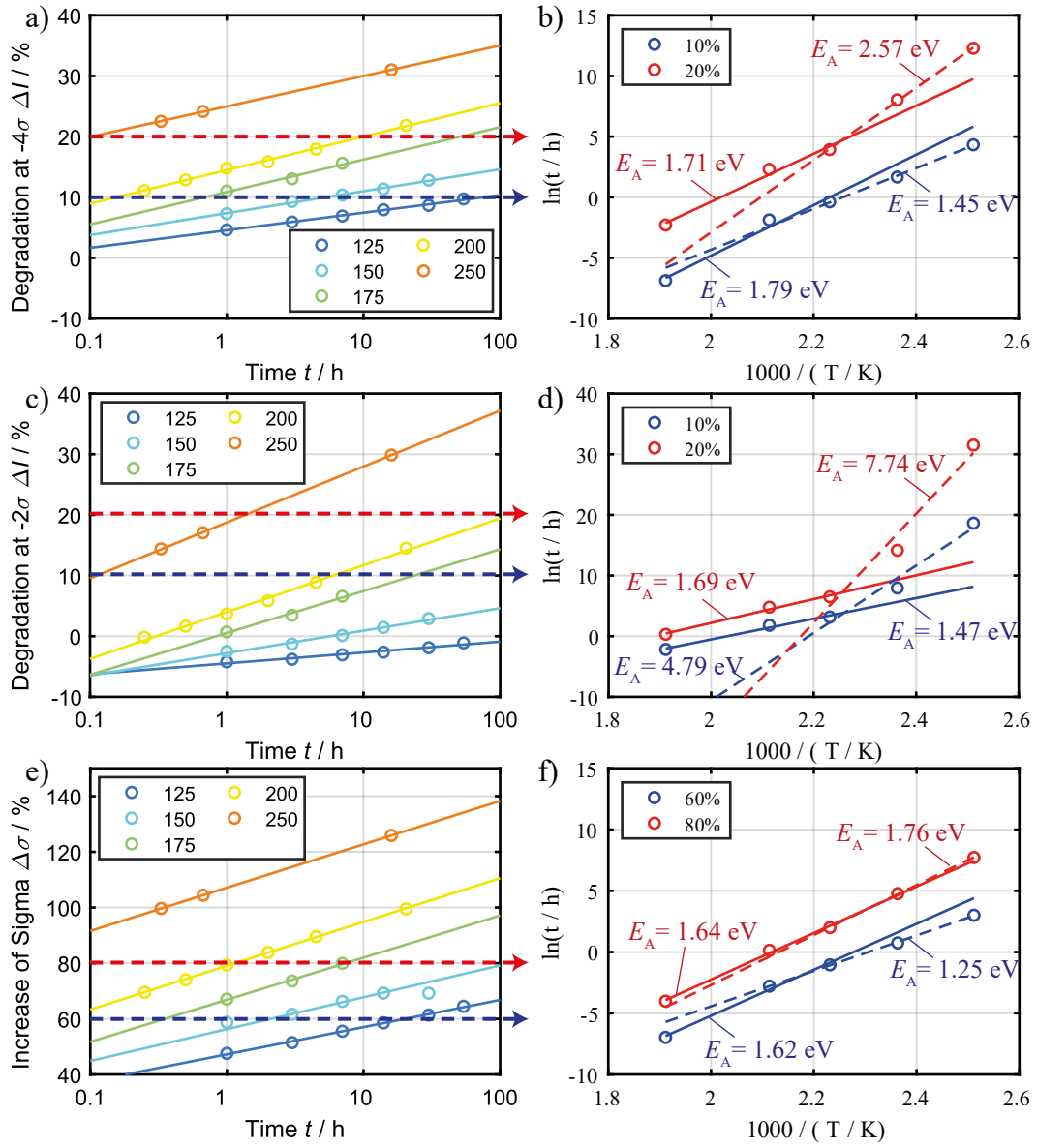

FIGURE 5.7: Determination of activation energies. a) Current change at -4σ in percent for different temperatures (legend in °C). b) Corresponding Arrhenius plot for 10 % and 20 % degradation. Different fits can be applied, resulting in different activation energies E_A. c) and d) equivalent to a) and b) for current at -2σ; At low temperatures (dashed in d)) the activation energy significantly deviates from the one at -4σ. Analogously, E_A is determined for the standard deviation σ in e) and f), which shows the correlation of tilting and degradation at low percentiles.

Regarding Arrhenius' law, lines corresponding to the same activation energy should be parallel on a logarithmic time scale. Therefore, figure 5.7, a) already suggests that the degradation follows different activation energies at higher temperatures compared to 125 °C and 150 °C. To determine the activation energy, the time when the current at -4σ degraded by 10 % or 20 % respectively is read from figure 5.7, a) and drawn into the Arrhenius plot in b). Here, the data is fitted linearly and the activation energy is read from the slope of these fits. As mentioned before, different activation energies can be determined for higher and lower temperature. Thus, two different fits are applied for 10 % or 20 % degradation each and the activation energies differ significantly. At higher temperatures the activation energy read at 10 % and 20 % degradation is comparable, whereas at lower temperatures a difference of more than 1 eV is observed. This shows that the activation energy not only depends on the bake temperature but may also change over time (from 10 % to 20 % degradation).
Another important factor is the percentile of the distribution for which the activation energy is extracted. In figure 5.7, c) and d) E_A is determined analogously for the degradation at -2σ. For higher temperatures the activation energy is again comparable to the preceding results but at lower temperatures unrealistic values of $E_A > 4$ eV are determined. This can be explained by the initial decrease of the read current which is more pronounced at shorter bake times and lower temperatures. The current decrease is revealed by negative values for the degradation in figure 5.7, c). Thus, it can be stated that the determination of activation energies via degradation of read current is affected by

- percentage of degradation (i.e. time)
- percentile of the distribution for which it is extracted.

Additionally, the underlying process may differ between temperature ranges, resulting in different activation energies at different bake temperatures. This is a major drawback if the activation energy is to be used to extrapolate high temperature retention towards lower operating temperatures.
This dissertation proposes to evaluate retention characteristics regarding the parameters of the intrinsic (log-normal) statistics observed in HfO_2 based ReRAM. As discussed before, the degradation of the programmed states can be described as tilt (σ) and shift (μ) of the read current distributions. Therefore, activation energies are determined for the change of the standard deviation σ in figure 5.7, e) and f). It can be seen that the difference between high and low temperature degradation is lower than observed before. Furthermore, the activation energies evaluated at 60 % and 80 % degradation are consistent with the high temperature E_A evaluated at -4σ and -2σ. This indicates that the tilting of the distribution is the major origin of the current

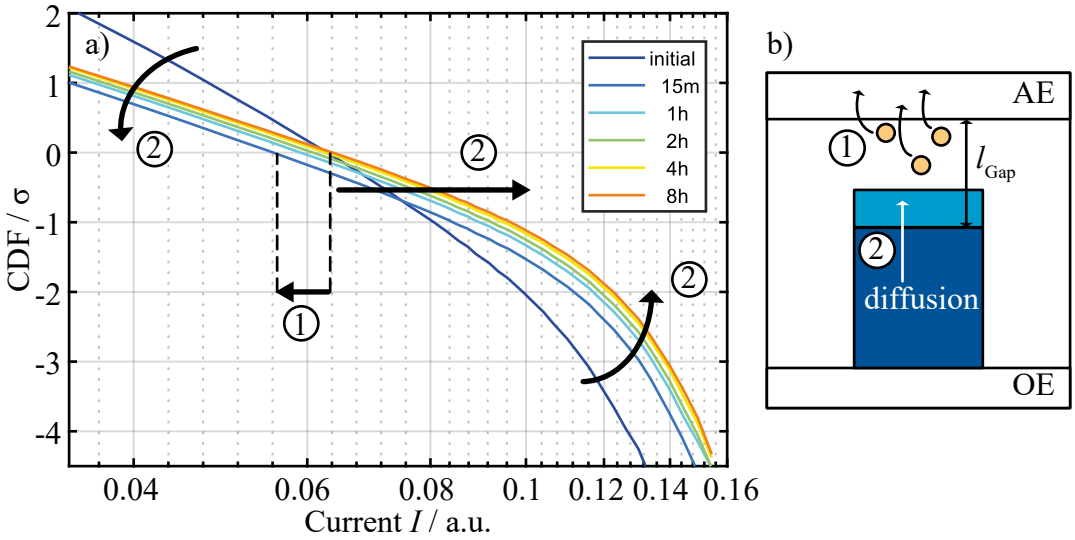

FIGURE 5.8: Proposed retention model. The initial decrease of the read current (1) is caused by recombination of $V_O^{\cdot\cdot}$ with oxygen from the active electrode (AE). A second process (2) causes the subsequent current increase and tilting/broadening of the distribution. (2) is considered to be the diffusion of $V_O^{\cdot\cdot}$ from the filament toward the gap at the AE. Reproduced with permission from S. Wiefels et al. [210], © 2020 IEEE.

degradation at these percentiles if $T \geq 175$ °C. The effect of the distribution shift (μ) is superimposed and is more apparent at lower temperatures. In conclusion, it can be stated that a reliable assessment of the activation energy for the processes involved in the HRS degradation is challenging. After all, the evaluation of the characteristic parameters of the distribution is the more promising approach to determine the retention at operating temperatures. The extracted activation energies in figure 5.7 are summarized in table 5.1.

TABLE 5.1: Activation energies for different criteria

E_A / eV	-4σ		-2σ		StDev σ	
	10 %	20 %	10 %	20 %	60 %	80 %
$T \geq 175$ °C	1.45	2.57	4.79	7.74	1.25	1.76
$T \leq 175$ °C	1.79	1.71	1.47	1.69	1.62	1.64

Interpretation

To explain the presented effects, the model illustrated by figure 5.8, b) is proposed. It assumes a cell with a conducting filament (blue) reaching from the ohmic electrode (OE) to a tunneling gap close to the active electrode (AE). The initial decrease of the median μ (figure 5.5, b)) can be explained by recombination of few remaining $V_O^{\cdot\cdot}$ in the gap close to the AE with oxygen from the active electrode. A lower concentration

of $V_O^{\cdot\cdot}$ should increase the resistance and thus shift the distribution towards lower read current. The depletion of $V_O^{\cdot\cdot}$ close enough to recombine or the depletion of available oxygen in the active electrode terminates this process labeled as ① in figure 5.8. In this model, a second process ② causes both the tilt of the distribution and the shift towards higher current, i.e. diffusion of $V_O^{\cdot\cdot}$ from the filament region towards the gap at the active electrode. A further discussion of this model is given in chapter 5.3.
With the extracted trends of the characteristic parameters of the distributions, it is possible to estimate log-normal distributions at different times and temperatures than covered in this study. This means that the derived understanding of the temperature impact on these parameters can be used to extrapolate the measured data to application temperatures and relevant retention time requirements. Since especially the current close to the read window is of importance and the observed distributions deviate from the ideal log-normal statistics at this end, the extrapolation is more challenging. Therefore, an empirical model which covers both the log-normal part of the distribution and the bending at higher currents, is developed and will be presented in the following section. Subsequently, chapter 5.3 will discuss the origin of the current limiting effect and deliver an explanation.

5.2 Empirical Model

As demonstrated in the previous section, the extraction of activation energies is not sufficient to estimate reliably if the device under test meets the retention requirements at operating temperature. Therefore, an alternative approach is presented in the following. Based on the model depicted by figure 5.8, an empirical tunneling model is developed and fitted to the experimental read current distributions. The extracted trend of the characteristic parameters μ and σ of the log-normal distribution is subsequently used to simulate read current distributions for different temperatures and bake times.
To isolate μ and σ as function of bake time and temperature, and to be able to extrapolate these parameters to values which are not covered by the experiments, two steps are required: At first, both parameters are fitted linearly regarding their respective time scale, i.e. $\mu \propto t^{1/4}$ and $\sigma \propto \ln(t)$, as depicted in figure 5.9, a) and b). Due to the limited temporal resolution, the initial decrease of μ can not be resolved and has to be excluded from the empirical model. The fit in figure 5.9, a) thus only considers data points after this shift. These fits give μ and σ as function of the bake time, but only for the four distinct temperatures covered in the experiment. To determine the general trend regarding the bake temperature, the slopes of the

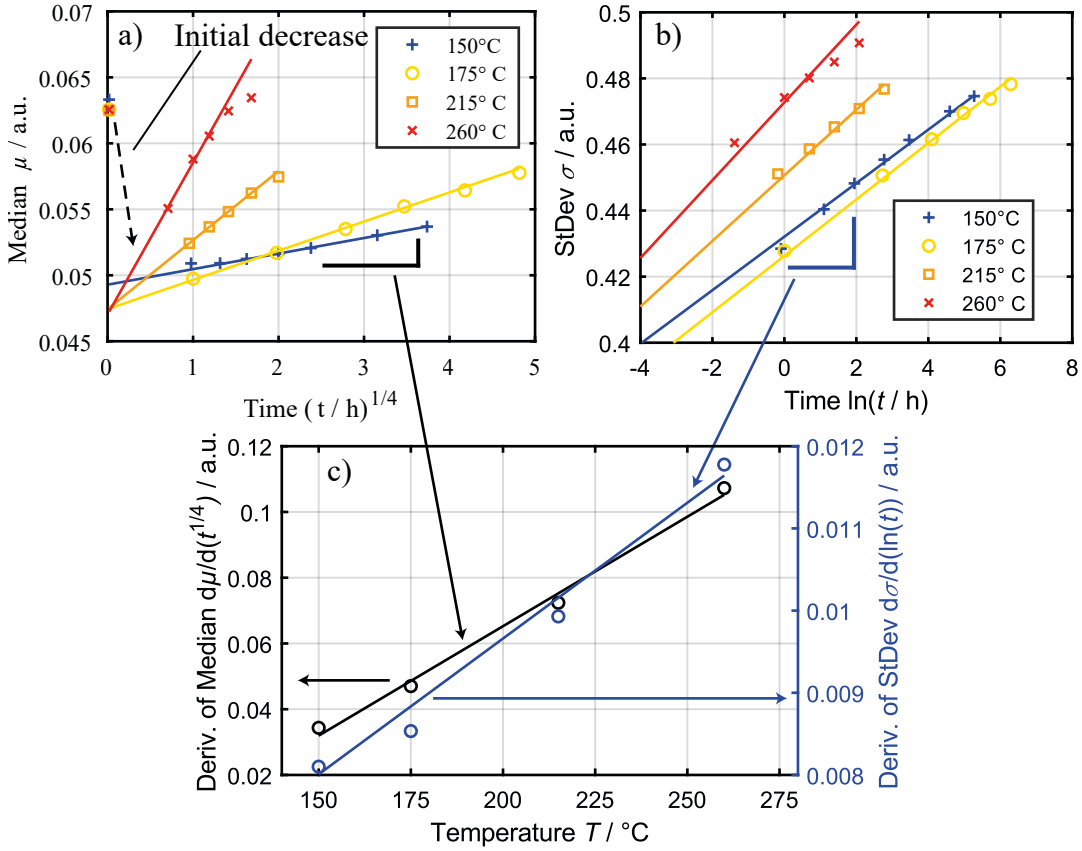

FIGURE 5.9: Extraction of the trend of μ and σ as function of bake time t and temperature T. a) μ is fitted linearly regarding $t^{1/4}$. The initial decrease is excluded. b) Linear fit of σ vs ln(t). c) The slopes of both fits are plotted as rate of change vs T. The linear fit provides μ and σ as f(t, T).

fits in figure 5.9, a) and b) are extracted and plotted versus the respective bake temperature in figure 5.9, c). This plot depicts the rate of change of μ and σ as function of the temperature. Another linear fit is added and results in the required correlation of μ and σ with time and temperature. This enables the estimation of the characteristic parameters for bake times and temperatures which are not covered by the experimentally obtained data. The limits of the potential to extrapolate will be discussed later.

5.2.1 Model V1

With the obtained correlation of μ and σ with t and T, the next step is to establish a procedure which translates these parameters into a read current distribution. Here, it is particularly challenging to match the bending or current limited regime. Based on the model in figure 5.8, a tunneling current across a gap is considered. Analogously

TABLE 5.2: Parameters of the empirical tunneling current

I_0	Fitting parameter	$6.7751 \cdot 10^{-5}$ a.u.
m^*	Electron tunneling mass	$9.10938 \cdot 10^{-31}$ kg
Φ	Tunneling barrier	0.75 V
V	Read voltage	0.2 V
A	Fitting parameter	$1.1757 \cdot 10^{-10}$ m
B	Fitting parameter	$6.2945 \cdot 10^{-12}$ m

to chapter 4.3.1, this current I_{Emp} can be expressed by

$$I_{\mathrm{Emp}} = I_0 \exp\left(-2d\hbar\sqrt{2m^*(e\Phi - eV)}\right), \tag{5.1}$$

with the reduced Planck´s constant $\hbar$ and the elementary charge e. The other parameters used in the equation are collected in table 5.2. The key variable in equation (5.1) is the length of the tunneling gap d. With the exponential dependence of I_{Emp} on d, a log-normal distribution is generated by a normally distributed tunneling gap d. This can be written as

$$d = d_0 + \Delta d \cdot n_{\mathrm{SND}}, \tag{5.2}$$

with the median gap length d_0 and a standard normally distributed random variable n_{SND} multiplied by the standard deviation Δd. By resolving equation (5.1) for $d = d_0$, the median gap can be directly calculated from the median current $I_{\mathrm{Emp}} = \mu$:

$$d_0 = \frac{\ln(I_0) - \ln(\mu)}{2\sqrt{2m^*(e\Phi - eV)}}. \tag{5.3}$$

The parameter Δd is fitted to the standard deviation of the current, via

$$\Delta d = A \cdot \sigma + B. \tag{5.4}$$

The fitting parameters A and B are given in table 5.2.
With this formulation and d_0 and Δd fitted to the experimentally determined μ and σ, the log-normal part of the read current distributions can be modeled. To generate a distribution of 2.5 M devices, the same amount of random numbers are drawn from the standard normal distribution and transferred into 2.5 M normally distributed tunneling gap values, using equation (5.2). With equation (5.1) this results in 2.5 M log-normally distributed read current values.
To introduce the observed current limiting effect, a minimum value of the gap length d

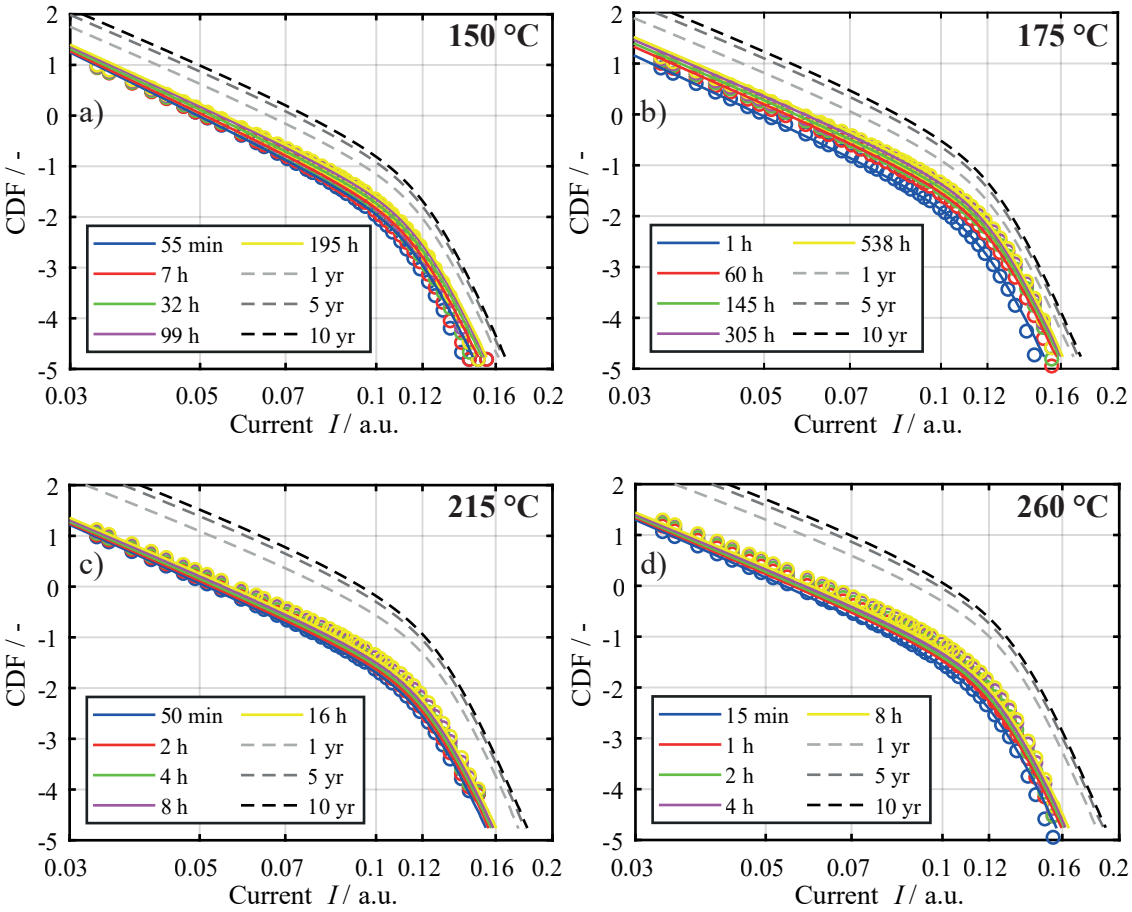

FIGURE 5.10: Results of the empirical tunneling model V1. Simulated distributions (lines) are given for the bake times covered by experimental data (circles) in the respective color. Additionally distributions for significantly longer bake times are simulated (dashed gray lines). For each bake temperature, the first read after the initial current decrease (blue) is evaluated for the characteristic parameters μ and σ. Distributions for later bake times are calculated based on the determined rate of change of μ and σ (c.f. figure 5.9). Reproduced with permission from S. Wiefels et al. [209], © 2020 IEEE.

is introduced as d_{min}, extending equation (5.2) to

$$d = d_0 + \Delta d \cdot n_{\text{SND}} \geq d_{\text{min}}. \tag{5.5}$$

Thus, during the generation of the gap distribution, all values of $d < d_{\text{min}}$ are discarded and generated again. If this minimal gap d_{min} is set to a fixed value the resulting current distribution would follow log-normal statistics until the respective current value for d_{min} and then be cut off sharply instead of the observed bending when approaching the current limit. Therefore, d_{min} is chosen to be normally distributed as well, reading

$$d_{\text{min}} = (d_{\text{min},0} + \Delta d) + \Delta d_{\text{min}} \cdot n_{\text{SND}}, \tag{5.6}$$

with its median $(d_{\mathrm{min},0} + \Delta d)$ and standard deviation Δd_{min}. It is found that $d_{\mathrm{min},0}$ and Δd_{min} can be interpreted as fitting parameters and kept constant. This means that the minimum gap d_{min} only depends on Δd which is directly correlated to σ after equation (5.4). Thus, the empirical distribution only has the two parameters μ and σ to fit the experimental distribution, including the limitation effect. With the extracted trend of these parameters shown above, empirical distributions can be generated as shown in figure 5.10. Plotted is the experimental data in circles and the simulated distributions as solid lines in the colors of the corresponding experimental bake times. Since the initial shift is not covered by the model, the read distribution after the first bake (blue) is evaluated regarding μ and σ. Based on their determined trend regarding time and temperature, μ and σ are calculated for the following bake times at the respective temperature and fed into the empirical tunneling model resulting in the presented distributions. It can be seen that the simulated distributions are in very good agreement with the experimental data. Additionally, μ and σ are calculated for significantly longer bake times and translated into read current distributions for theoretical 1 year, 5 years and 10 years at the respective temperature. These extrapolated distributions are represented by dashed lines in figure 5.10. This allows for an estimation of the complete distribution, and in particular the read window at the end of the required retention time.

In figure 5.10, simulated read current distributions are only given for bake temperatures which are covered by the set of experimental data. Though, the aim of the empirical model is to estimate the retention characteristics of the device under test for operating temperatures. In order to do this, the rate of change of μ and σ can be calculated for the desired temperature using the fits in figure 5.9, c). The same data is plotted again in figure 5.11 to emphasize a challenge deriving from the extrapolation of these fits. If the desired temperature lies within the tested temperature range (solid lines, 150 °C to 260 °C), the calculated rate of change has a high probability to be correct. Considering the desired lower bake temperatures, it is not proven that the estimated trend will continue. Regarding the rate of change of the standard deviation (blue), the extrapolation towards room temperature (25 °C) is plausible. But for the median μ, the rate of change becomes negative below 100 °C. This is generally possible and could be connected to the initial current decrease observed after programming. Though, at some point this trend has to become implausible: Considering very low temperatures, the current distribution would be supposed to accelerate towards lower current, although temperature activated processes as diffusion of defects should be inhibited. Therefore, and due to the limited experimental data, an alternative extrapolation is suggested. Below 150 °C the rate of change is considered to follow a straight line towards the absolute lower limit of 0 K. Thus, the

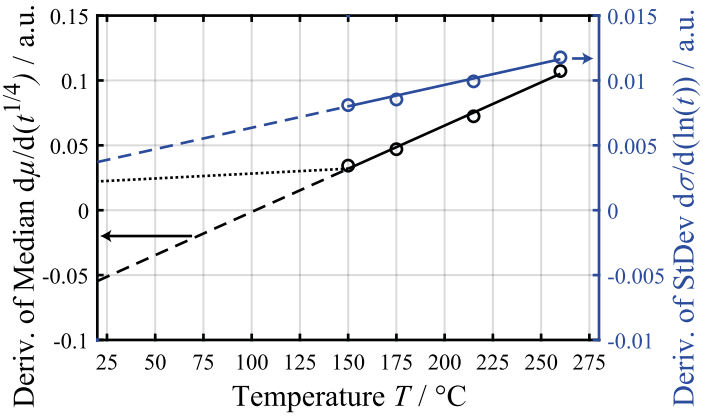

FIGURE 5.11: Rate of change of μ and σ, extracted from experimental data. Linear extrapolations towards room temperature are added in dashed lines. The extrapolated rate for μ becomes negative at approx. 100 °C. An alternative extrapolation towards 0 K is given by dotted line.

resulting values for operating temperatures between room temperature and 150 °C are supposedly estimated higher than in reality. But, to ensure a certain read window after the required retention time, this gives a more reliable estimation. Resulting extrapolations for the retention at room temperature are calculated using an updated version (V2) of the empirical model which will be presented in the following section.

5.2.2 Model V2

It was demonstrated above that the empirical model is well suited to describe the retention of the industrial HfO_2 VCM devices. However, it relies on the calculation of normally distributed random variables which becomes a huge computational burden if large sets of devices are to be simulated. To extrapolate the critical current at very low percentiles (e.g. $-6\sigma \approxeq 1$ GBit) an analytical formalism of the model is required. The measured HRS distributions for the retention experiment at 175 °C are plotted again in figure 5.12 (circles). It can be clearly seen that the distributions are comprised by two linear slopes connected by a bending at approx. 0.12 (a.u.). The decreasing line y_1 describes the discussed log-normal part of the distribution and is therefore determined by

$$y_1 = \frac{1}{\Delta d} \cdot (d_2 - d_0), \text{with} \tag{5.7}$$

$$d_2 \in [7, 11] \cdot 10^{-10}\text{m}. \tag{5.8}$$

Equal to model V1, the parameters Δd and d_0 are extracted from the experimental distributions via equations (5.3) and (5.4). The second line y_2 is found to follow the equation

$$y_2 = C \cdot (d_2 - d_{\text{min},2} + \Delta d), \tag{5.9}$$

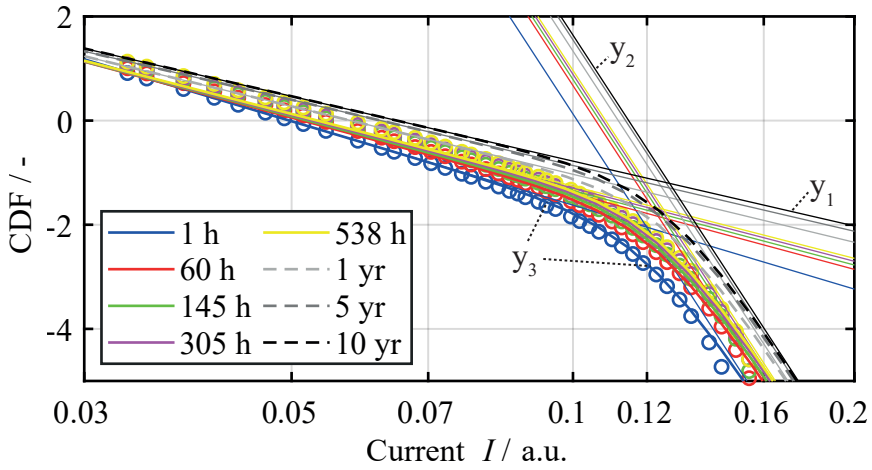

FIGURE 5.12: Experimental retention data for 175 °C (circles). The log-normal part of the distribution is modeled by y_1. The current limiting regime is modeled by y_2. The combination of y_1 and y_2 yields the simulated distributions y_3. Exemplary extrapolations towards 1 yr, 5 yrs and 10 yrs are depicted by dashed lines.

TABLE 5.3: Additional parameters of the model V2

C	Fitting parameter	$9.0909 \cdot 10^{10}$ m^{-1}
$d_{\mathrm{min},2}$	Fitting parameter	$9.13 \cdot 10^{-10}$ m

with the fitting parameters C and $d_{\mathrm{min},2}$, given in table 5.3. Consistently with model V1, the shape of the distribution in the current limiting regime only depends on Δd, representing the reciprocal slope of the main, log-normal part.
The actual distribution y_3 is constructed by the combination of y_1 and y_2 via

$$y_3 = -\ln(\exp(-y_1) + \exp(-y_2)). \tag{5.10}$$

As shown in figure 5.12, the simulated line y_3 matches the experimental data very well. The computational burden is reduced by several orders of magnitude by using equations (5.7)-(5.10) instead of the massive generation of random numbers.
Finally, the empirical model (V2) is used to calculate read current distributions for the degradation at lower temperatures to estimate the retention characteristics at operating temperatures. Since the empirical model does not cover the initial current decrease, the starting point of this extrapolations must be a distribution measured after the initial decrease. As depicted by blue circles in figure 5.13, the experimental distribution after 1 hour at 175 °C is evaluated regarding the characteristic parameters of the log-normal distribution μ and σ. Subsequently, these parameters are extrapolated using the rates of change obtained by figure 5.11. Here, the extrapolation towards -273.15 °C (0 K, dotted line) is chosen for the rate of change of μ as discussed in the previous section. The extrapolated distributions are calculated for 0 °C, 25 °C,

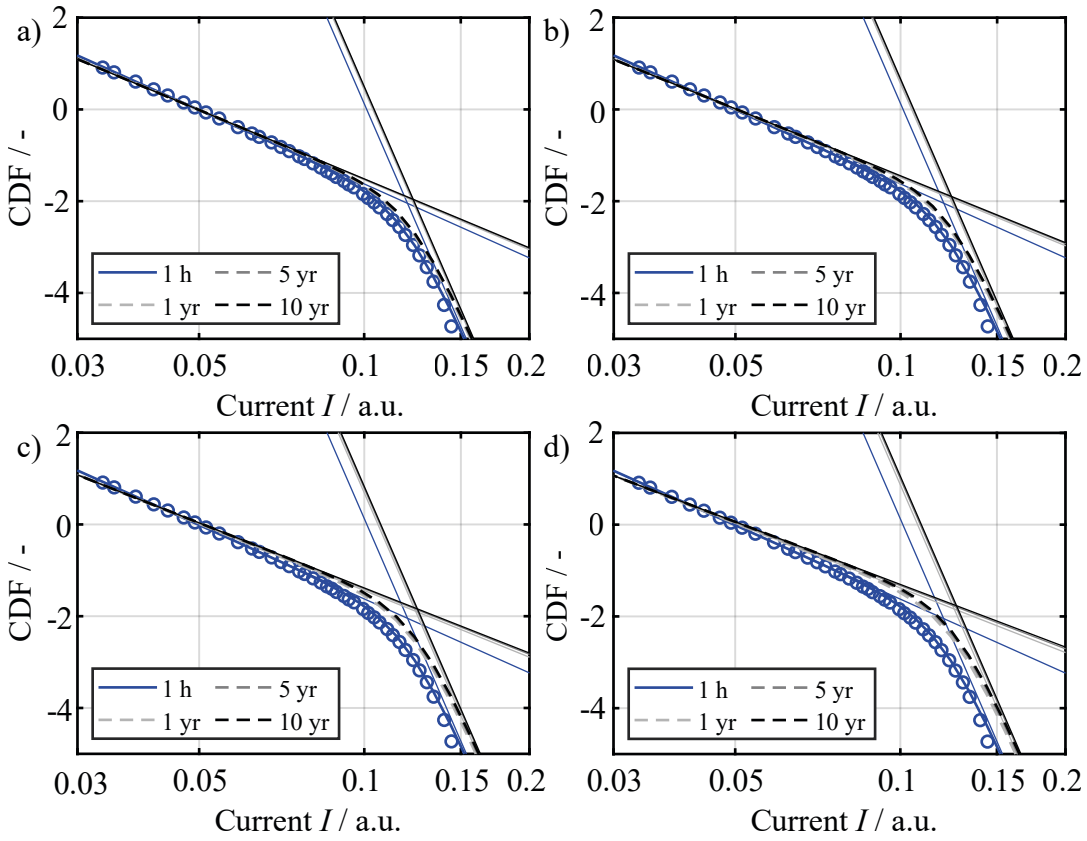

FIGURE 5.13: Simulated read current distributions for the extrapolated retention characteristics at a) 0 °C, b) 25 °C, c) 50 °C, d) 85 °C. In each case, the experimental data for 1 hour at 175 °C (after initial decrease) is shown in circles, evaluated regarding μ and σ, and used as starting point of the extrapolation. The straight lines depict the the components y_1 and y_2 of the modeled distribution (c.f. fig. 5.12, eq. (5.7), (5.9)).

50 °C and 85 °C to cover a wide range of possible operating temperatures and shown in figure 5.13. At the lowest temperature calculated (fig. 5.13, a)), the distribution is extremely stable. Although the rate of change of μ is estimated higher than expected (extrapolation towards 0 K), the shift in μ is close to zero, even after 10 years, as can be seen from the current at 0σ. For lower percentiles, close to the read window, a slight increase of the read current is observed which can be attributed to tilting of the distribution, i.e. the increase in σ. For 25 °C and 50 °C in figure 5.13, b) and c) the shift of the current median is still negligible. It can be seen that the increase in σ rises with temperature, directly impacting the critical current region close to the read window. This trend continues with the simulation for 85 °C in figure 5.13, d). Here, a slight impact on the current median is observed which superimposes the current change at low percentiles. Though, the tilting of the distribution still seems to be the predominant factor regarding the degradation of the read window. On the one hand, this is beneficial for the empirical model. With the discussed doubts regarding the extrapolation of the rate of change in μ, the fact that μ hardly impacts the retention

characteristics significantly increases the trustworthiness of the model. On the other hand, it shows that the major challenge in optimizing the retention of a material under test is to reduce the width of the distribution and to suppress its broadening (increase in σ) as far as possible.
All in all, it can be stated that the empirical model is very well suited to fit the experimental read current distributions, as well as the observed retention characteristics. With the presented extrapolations of the characteristic parameters μ and σ, the model allows to estimate the degradation of a distribution (excluding the initial current decrease), for the required retention time at the given operating temperature.
The advantage of the second version of the empirical model is the massive reduction of the computational burden. This enables to estimate the most critical current at very low percentiles (close to the read window) for various times and temperatures. In figure 5.14, a) the simulated trend of the current at -5σ is shown for the measured temperatures 150 °C, 175 °C, 215 °C and 260 °C. In each case, the starting point is the read after the first bake, i.e. after the initial decrease of the read current. The model gives a clear trend of the critical current, increasing linearly on a logarithmic time scale. With the corresponding current at the lower end of the LRS distribution (not shown), the read window can be directly estimated for any retention time. figure 5.14, b) shows the trend of the standard deviation σ for the simulated distribution underlying the values in a). The comparison underlines again that the change in σ is the primary origin of the current change at low percentiles.
Using the empirical model, the extrapolation of the critical current is extended to various temperatures ranging from 0 °C to 250 °C in figure 5.14, c). Here, the initial value for all calculations is drawn from the retention experiment at 175 °C. Again, the current of the first read after the initial current decrease is chosen (i.e. the initial current of the yellow line in figure 5.14, a)). Using the analytical version (V2) of the empirical model, numerous distributions are simulated in the presented temperature range and for retention times up to 20 years. Analogously, the current at even lower percentiles can be calculated to account for even larger memory arrays. figure 5.14, d) shows the simulated trend of the current at -6σ which would represent the worst bit in approx. 1 GBit. Thus, the empirical model provides a simple and (computationally) cost-effective way to estimate the retention of very large ReRAM arrays. As discussed above (c.f. section 5.1.2), the determination of an activation energy for the occurring degradation processes is challenging. Multiple underlying processes and also intrinsic statistics of ReRAM devices impede reliable extrapolations with this method. The presented empirical tunneling model represents an alternative approach, accounting for the whole current distribution with its underlying statistics. By extrapolating the trends of the characteristic parameters of the intrinsic log-normal

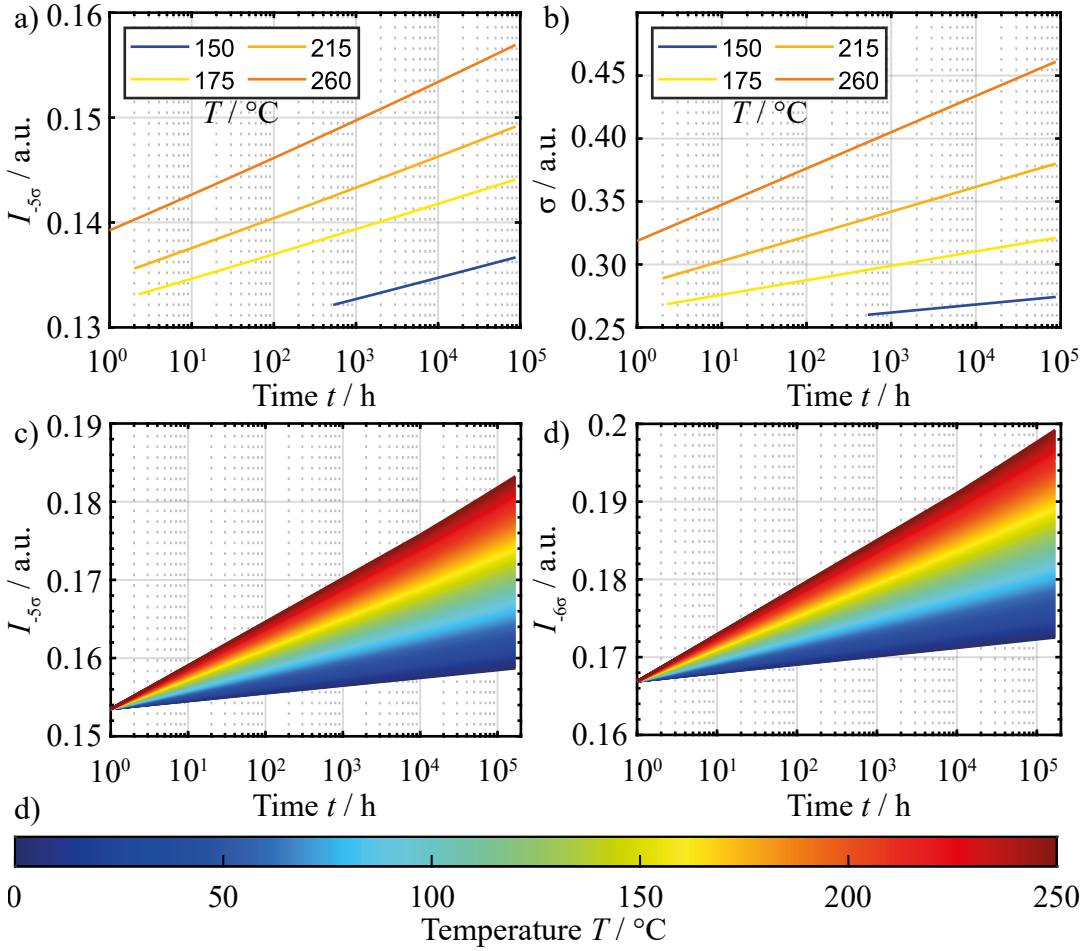

FIGURE 5.14: a) Read current close to the read window (-5σ) extrapolated for 150 °C, 175 °C, 215 °C and 260 °C up to 10 years. The critical current increases linearly with log(t). b) Trend of the standard deviation σ, corresponding to the data in a). It becomes clear that the increase in σ is the predominant origin of the trend in a). c) Current at -5σ calculated for various temperatures from 0 °C to 250 °C. d) Trend of the current at -6σ analogously. e) Colorbar for the temperatures in c) and d).

statistics, reasonable estimations even for the weakest cell in a large array can be made.

It may be noted that despite physically reasonable parameters are chosen, the presented model only gives an empirical description of the read current. To approach a physical understanding of the underlying processes a statistical retention model is presented in the following section.

5.3 Statistical Model

As demonstrated in the previous chapters, the internal statistics of VCM ReRAM devices are one of their most important characteristics and a key challenge for the reliability of large sets of cells. Therefore, a statistics based simulation model is

developed which is presented and discussed in chapter 4.3.2. In the following, this model is extended to a retention model using two different approaches. On the one hand, the static model, presented in chapter 4.3.2, is exploited by changing the boundary conditions of the filament to simulate degradation effects. On the other hand, a dynamic model is derived which also includes the KMC algorithm. Here, the division of the cell into regions exhibiting different activation energies for the ion migration (diffusion) will explain the parallel occurrence of variability and high temperature (or long time) degradation [208].

5.3.1 Static Model

In chapter 5.1.1 and 5.1.2 the degradation effects on a log-normal read current distribution are identified as

(i) shifting of the current median μ, and

(ii) tilting (i.e. broadening) characterized by a change of the standard deviation σ.

As a first approach to explain the observed retention characteristics, the static variability model (c.f. chapter 4.3.2) is exploited as a retention model. As discussed before, the general concept of the variability model is a random configuration of defects within a filamentary region ("box"). It is demonstrated in chapter 4.3.2 that experimentally observed normal and log-normal read current distributions can be modeled by assembling the read current of a respective number of configurations, generated by random placement of defects within the right boundary conditions. These boundary conditions can be summarized as

- geometric measures of the filament, and
- the number of vacancies per filament.

Because the general statistics (i.e. the shape) of the read current distributions stay intact during retention experiments, the static variability model is able to describe the initial state as well as the degraded state after any of the presented bake experiments. If the general concept of random defect configurations still applies, it is standing to reason that the boundary conditions must have changed during bake. Therefore, variations of the boundary conditions are considered as degradation mechanisms and characterized for their impact on the read current distribution.
The investigated variations, summarized in figure 5.15, are

(a) increasing width of the filament, i.e. increasing filament area,

(b) increasing length of the filament, i.e. decreasing gap at the active electrode, and

(c) decreasing number of defects via exchange with the electrodes.

Process (a) and (b) cover the change of the filament geometry. It is generally considered that the filament contains a higher number of defects than its surrounding. The resulting concentration gradient determines the expected change of the filament box. Therefore, in both cases, the assumed direction of change is the one towards increasing filament volume.
Process (c) represents the change of the defect concentration. Here, the number of defects can either decrease by recombination of oxygen vacancies with oxygen from one of the electrodes or increase by generation of vacancies at the electrodes. A generation of oxygen vacancies as anti-Frenkel pairs in the bulk is reported to be extremely unlikely [182]. The expected direction of process (c) depends on the chemical properties of the respective oxide and electrode metal. As introduced in chapter 2.2.3 and discussed in chapter 6, the ZrO_2/Ta stack should exhibit a high defect formation energy. Therefore, it is assumed that the industrial HfO_2 cells as well comprise a defect formation energy sufficiently high to prevent a spontaneous generation of oxygen vacancies. Thus, a decrease of the number of defects is considered more likely as degradation mechanism than the opposite [208].
At first, the influence of an increasing filament area (process (a)) is investigated. Following the simulation flow in figure 4.12 b), read current distributions are generated with 30 defects, placed randomly into the filament volume. As explained before, each defect configuration is represented by one current value in the distribution. To simulate a large set of VCM cells, 1000 random configurations are generated for each current distribution. To achieve log-normal distributions, the gap length is set randomly within 0.75 nm and 2.75 nm. It may be noted that the impact of the

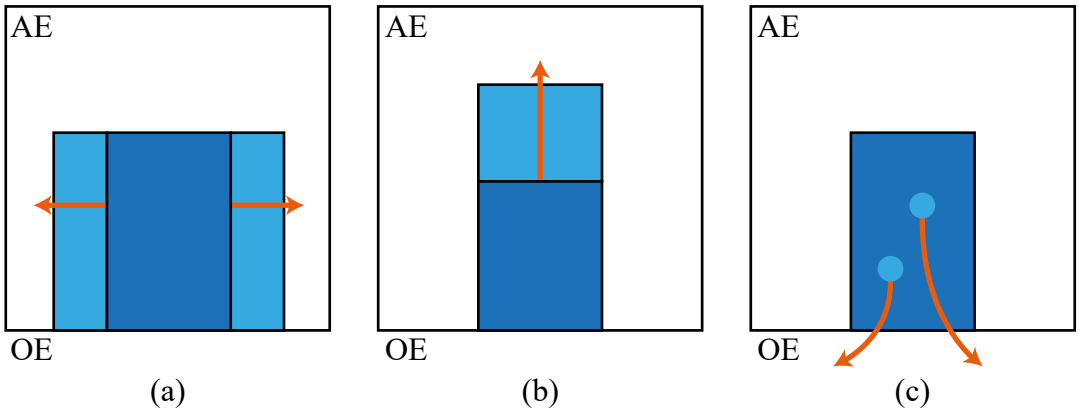

FIGURE 5.15: Schematic of the degradation effects applied to the boundary conditions of the static variability model. (a) Broadening of the filament, i.e. increase of the filament area A_{fil}. (b) Increase of the filament length, i.e. decrease of the gap l_{gap} between gap and active electrode (AE). (c) Decrease of the number of defects via recombination of oxygen vacancies with AE or ohmic electrode (OE). Adapted from [208].

individual gap length is not resolved in this simulation, but will be investigated later. Here, a certain distribution of gap lengths over all simulated cells is achieved which helps to accomplish log-normal statistics. Within the simulated distributions, the base A_{fil} of the filament is varied. Figure 5.16, a) shows exemplary configurations drawn from the simulation model with filament areas of $(2.5 \cdot 2.5)$ nm^2, $(3 \cdot 3)$ nm^2 and $(4 \cdot 4)$ nm^2. The resulting read current distributions are shown in figure 5.16, b) with a linear current scale. Similar to the experimental results shown in figure 5.3, a) the simulated data seems to develop two sub distributions. This effect is especially pronounced with smaller filament areas. Analogously to the experimental data, the simulated distributions are plotted on a logarithmic current scale in figure 5.16, c). This reveals on the one hand the expected log-normal statistics. On the other hand, the simulated distributions show the same current limiting effect as observed with the experimental results. Figure 5.16, e) reprints the experimental retention data for a 175 °C bake from figure 5.4, b). Regardless of the variation of the filament area, it can be seen that the simulated distributions are qualitatively in very good agreement with the experimental data.

Furthermore, the statistical simulation gives a possible explanation for the observed current limiting effect: As explained, the log-normal statistics are achieved by random defect configurations (and additional variation of the gap length) within given boundaries. At the same time, these boundaries limit the number of possible defect configurations. Thus, at some point, the "optimal" configuration (w.r.t. high read current) must be reached and therefore, the log-normal trend cannot continue endlessly. Depending on the total number of possible configurations, determined by the boundary conditions, the bending of the distributions occurs at higher or lower percentiles. This is observed in figure 5.16, c). With decreasing filament volume, a higher number of cells is affected by the current limit. It may be noted that the current value of this limiting effect might suggest the opposite trend: The cells with greater volume seem to exhibit a lower current limit. But, with trap assisted tunneling as conduction mechanism, a lower current is generally expected for broader filaments. This is explained by the higher mean distance between individual defects resulting in lower transition probabilities. The absolute limit of the read current should be independent of the outer boundaries of the filament area because the configuration with the highest conduction is expected to be a very densely packed chain of defects. However, the larger filament volume provides more possible configurations so that a lower percentage of cells will be close to the absolute current limit.

Equivalent to the analysis of the experimental retention data in chapter 5.1.2, the characteristic parameters μ and σ are extracted from the log-normal part of the simulated distributions and depicted in figure 5.16, d) and f). It can be seen that μ as

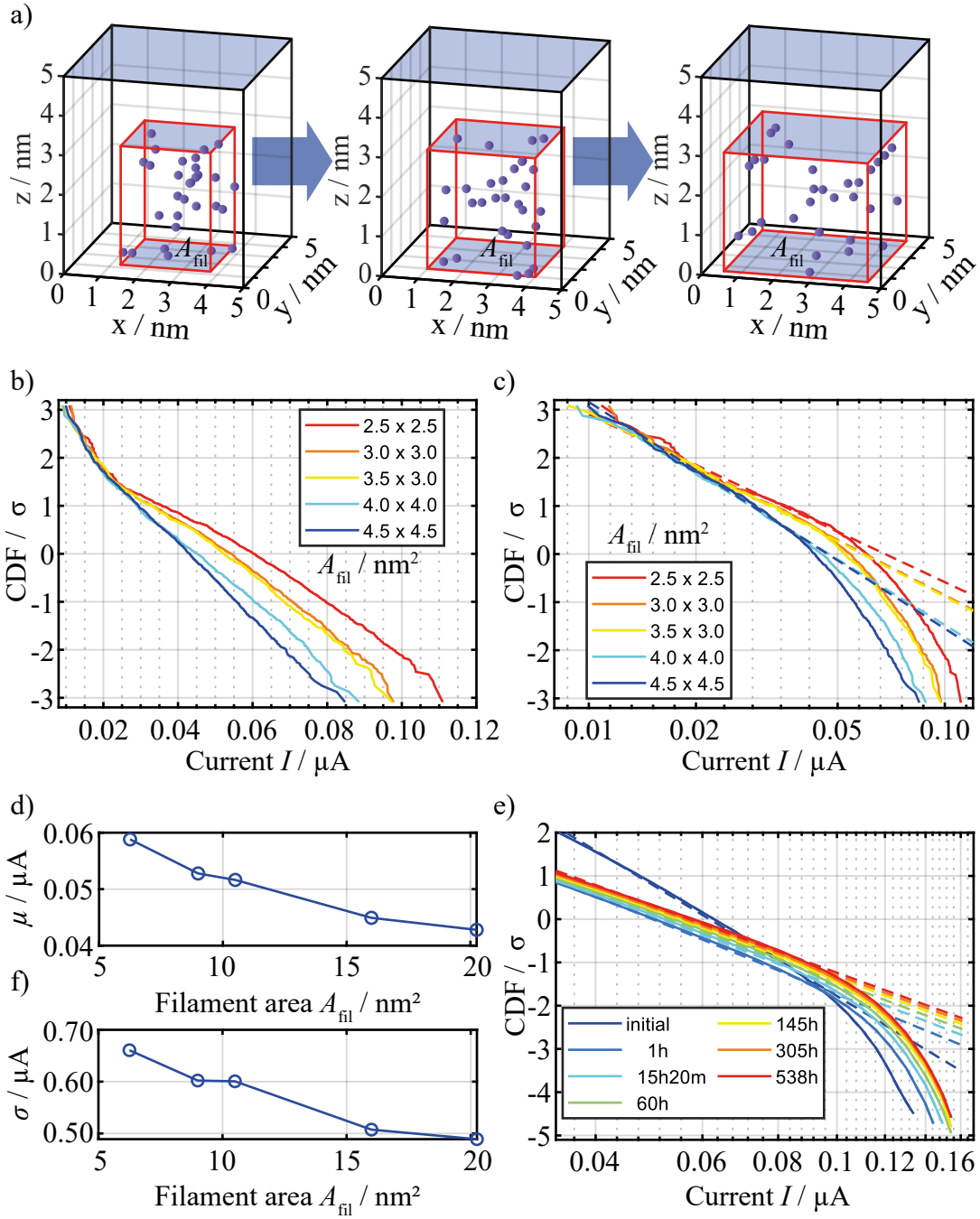

FIGURE 5.16: Statistical simulation with variation of the filament are A_{fil}. To achieve log-normal statistics, the gap length l_{gap} is set randomly between 0.75 nm and 2.75 nm. A_{fil} is varied between distributions but constant for all cells within one distribution. a) Exemplary defect configurations for filament areas of $(2.5 \cdot 2.5)$ nm^2, $(3 \cdot 3)$ nm^2 and $(4 \cdot 4)$ nm^2. b) Resulting read current distributions for different A_{fil} on linear current scale. c) The logarithmic current scale reveals log-normal statistics as highlighted by dashed lines, as well as the experimentally observed current limiting effect at the right hand side. d) Characteristic parameters μ and σ from the log-normal fit in c). Both parameters decrease with increasing A_{fil}. e) Measured distributions for a bake experiment at 175 °C for comparison. Partially reproduced with permission from S. Wiefels et al. [210], © 2020 IEEE.

well as σ decrease significantly with increasing filament area. As mentioned above, a decrease of the current median is expected because of the higher mean distance between the defects, resulting in a lower TAT current. Compared to the measured trend of μ during bake experiments, a significant decrease only fits to the trend during the first hour of bake. Here, an initial current decrease was observed followed by monotonously increasing current median. Therefore a broadening of the filament comes into account as explanation of the initial current decrease. But, considering that this process seems to be self limiting and is followed by the opposite trend, a broadening of the filament becomes unlikely. If the width of the filament increases due to lateral diffusion of defects, there should be no reason for this process to be limited, unless the outer boundaries of the cell are reached.
Regarding the trend of the standard deviation σ, the simulated decrease contradicts the expectations. A greater filament volume should result in a higher number of possible configurations and therefore a broader distribution. The deviation from the expected trend might be explained by the strong influence of the current limiting effect. In the simulated distributions, the bending starts around the current median (0σ). Compared to the measurement, where the bending is initiated at approx. -1.5σ, the standard deviation will be much more affected by the bending in the simulation. Though, this first simulation with the static version of the statistical retention model provides read current distributions in good agreement with the presented experiments. This includes the log-normal statistics, as well as the current limiting effect. The general concept of the statistical model furthermore explains this effect by the limited number of possible random defect configurations.
Due to the discussed problems of this simulation in extracting the trend of μ and σ regarding a broadening of the filament, a second approach is presented in figure 5.17. Here, the length of the filament is kept constant at $l_{\mathrm{gap}} = 2.75$ nm. The number of defects is reduced to 20 and the filament area is varied between $(1.5 \cdot 1.5)$ nm^2 and $(4.5 \cdot 4.5)$ nm^2. The resulting read current distributions are shown in figure 5.17, a). As expected with a constant gap, the data is normally distributed. Due to the reduced number of vacancies the overall current is lower and the current limiting effect is not observed. Again, the characteristic parameters are extracted and plotted in figure 5.17, b). The median μ of the read current follows the same decreasing trend as observed before, which can again be attributed to the higher mean distance of the defects in a larger volume. In contrast to the usual terminology of this work, the standard deviation σ in figure 5.17, b) represents the standard deviation of the linear data (usually the standard deviation of the logarithmic current is intended), because it describes the reciprocal slope of the normal distribution. In contrast to the previous simulation, σ exhibits no decreasing trend. Proceeding from the smallest

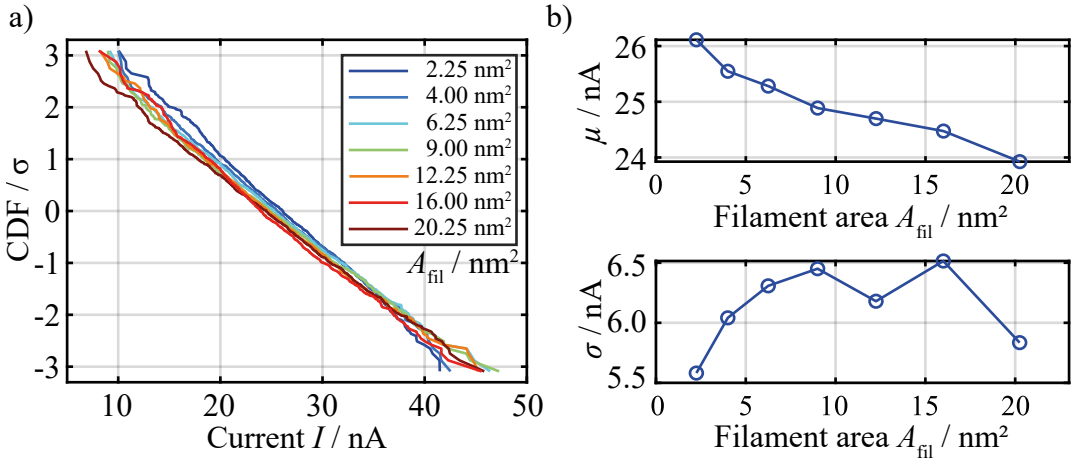

FIGURE 5.17: a) Simulated read current distributions for 20 defects and a fixed gap length of $l_{gap} = 2.75$ nm. Distributions are simulated for varying filament area, ranging from $(1.5 \cdot 1.5)$ nm^2 to $(4.5 \cdot 4.5)$ nm^2. b) Characteristic parameters extracted from the distributions. μ decreases with A_{fil} as expected. No clear trend can be determined for σ. Note that σ here is the standard deviation of the normal distribution (i.e. linear current).

area, it seems that σ rather increases with A_{fil}. Though, no clear trend can be determined. This means that a broadening of the filament does not provide a substantial explanation for the experimentally observed retention characteristics.

Secondly, the impact of a defect diffusion towards the active electrode, resulting in a reduced gap length l_{gap} is investigated. The simulation setup is illustrated by exemplary defect configurations in figure 5.18, a). Here, the base of the filament is kept constant at $A_{fil} = (2 \cdot 2)$ nm^2 and 30 defects are placed randomly in the filament volume. The length of the gap is varied between 0.75 nm and 3.75 nm. The resulting read current distributions are shown in figure 5.18, b). Because the gap length is constant within each individual distribution, the lines are linear on the linear current scale, indicating normally distributed read current. With decreasing tunneling gap, a clear trend is observed. As expected, the distributions shift strongly towards higher current, if the tunneling gap is reduced. Additionally, they tilt towards broader distributions and higher standard deviation σ.

However, at low values of $l_{gap} < 1.25$ nm the impact of the gap length on the read current seems to perish. This phenomenon might be explained by two aspects. On the one hand, the higher mean distance between defects with increasing filament volume could come into account as competing mechanism to the smaller tunneling gap. On the other hand, the tunneling process at the active electrode is affected by the Fermi distribution of the respective electrode metal. If the defect is too close to the electrode, there might be less free states in the electrode (or vice versa less free electrons) resulting in a reduced tunneling probability. Further simulations with the statistic model revealed that the latter is more probable as origin of this saturation.

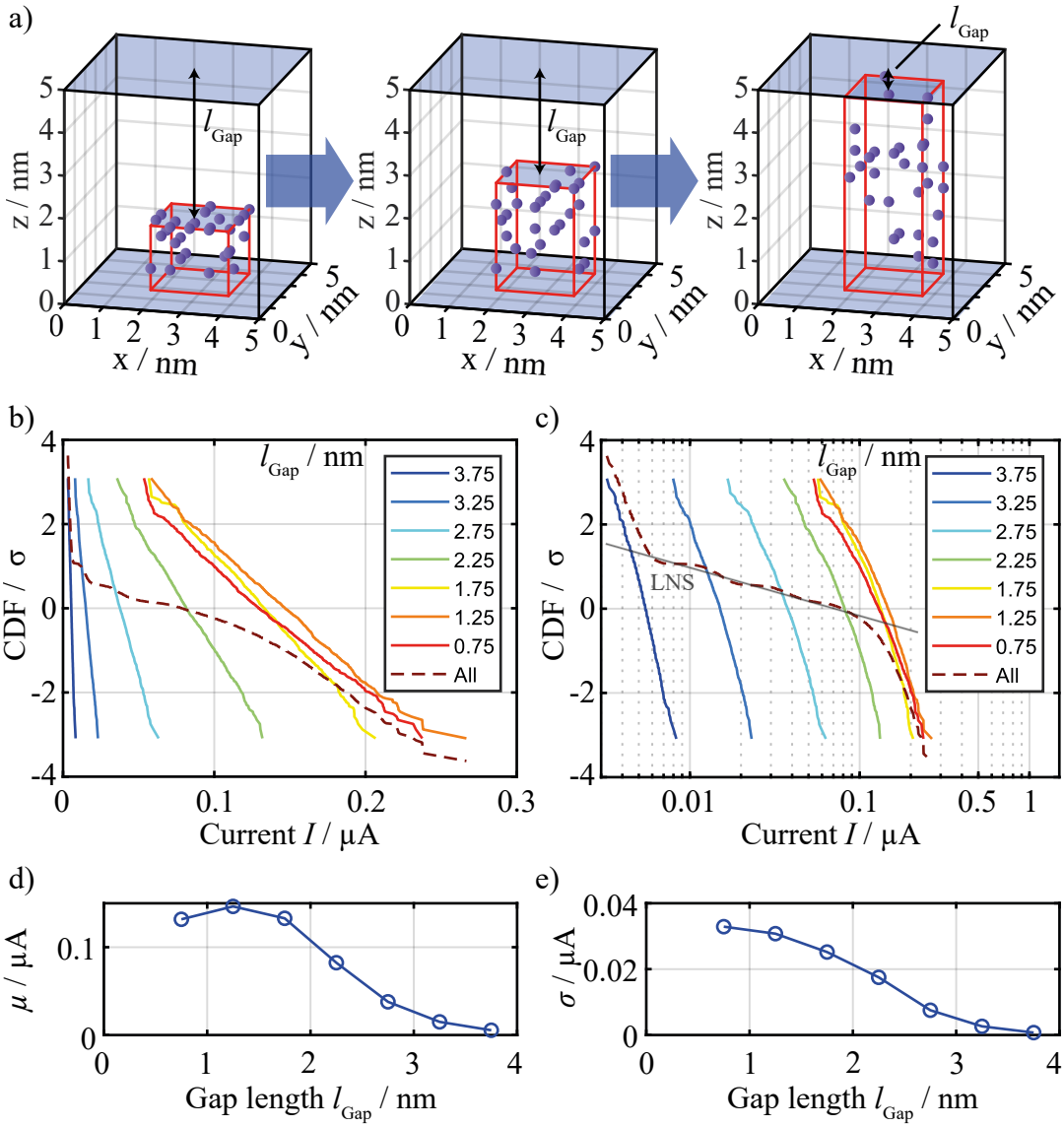

FIGURE 5.18: Static simulation for 30 defects with variation of the gap length l_{gap}. a) Exemplary configurations of defects with decreasing gap. b) Normally distributed read current distributions. A decreasing gap results in increased read current and tilting of the distribution towards higher σ. Note that σ here is the standard deviation of the normal distribution, accounting for linear current. The combined distribution for all gaps is given by the dashed line. c) data from b) on logarithmic current scale, revealing log-normal statistics (LNS) for the combined distribution. d) and e) Extracted trend of μ and σ. Both parameters increase with decreasing gap. At 1.25 nm the trend saturates to a plateau. Partially reproduced with permission from S. Wiefels et al. [210], © 2020 IEEE.

Furthermore, the normal distributions with different gaps are combined to the dashed line, representing a distribution with equally distributed tunneling gaps. On the logarithmic scale in figure 5.18, c) the main part of the combined distribution is observed to be log-normally distributed. At both ends of this distribution a bending or current limiting effect is observed. Here, the line approaches the normal distribution with the largest or smallest gap respectively. This underlines that the boundaries of the simulated system (here the borders of the gap) limit the intrinsic log-normal statistics. Again, the characteristic parameters of the distributions are plotted versus the tunneling gap in figure 5.18, d) and e). As explained, the median increases exponentially with reduction of the gap until the discussed limit at 1.25 nm is reached. At the same time, the standard deviation σ of the normal distributions (linear current) increases as well. At very small gaps σ also approaches a plateau. Compared to the experimental results, these trends fit very well the dominating degradation effects, being the increase in μ and σ over time (excluding the initial decrease in μ). Therefore, the diffusion of defects from the filament region towards the active electrode, resulting in a reduced gap, is considered as very likely mechanism of the high temperature or long term degradation.

To extend the understanding of this mechanism, the simulation is repeated with a significantly increased number of defects. Figure 5.19, a) illustrates random configurations of 200 defects with variation of the gap. It can be seen in the first frame of figure 5.19, a) that this high number of defects is very densely packed for large gaps. This leads to a very small number of possible configurations which results in a very steep read current distribution as observed in figure 5.19, b). Here, the same trend is observed for the normal distributions for different gaps, i.e. an increase of current and standard deviation with decreasing tunneling gap. The combined distribution is analogously given as dashed line. In comparison to the previous result, this combined set approaches a normal distribution as it is typically observed in LRS. Considering the higher number of defects and higher overall read current this transition towards normal statistics is very consistent with the experimentally observed transition from HRS to LRS. Though, the presented distribution with varying gaps is still considered a log-normal distribution as shown in figure 5.19, c). Furthermore, the combined distribution exhibits more discrete steps between the underlying normal distributions. This accounts for the fewer possible configurations due to the high number of defects in the filament volume.

The extracted trends of μ and σ are again shown in figure 5.19, d) and e). Despite the significantly increased number of defects, the trend is equal to the previous simulation, including the plateau below 1.25 nm. Thus, it can be concluded that a decrease of the tunneling gap is very likely to be the mechanism underlying the

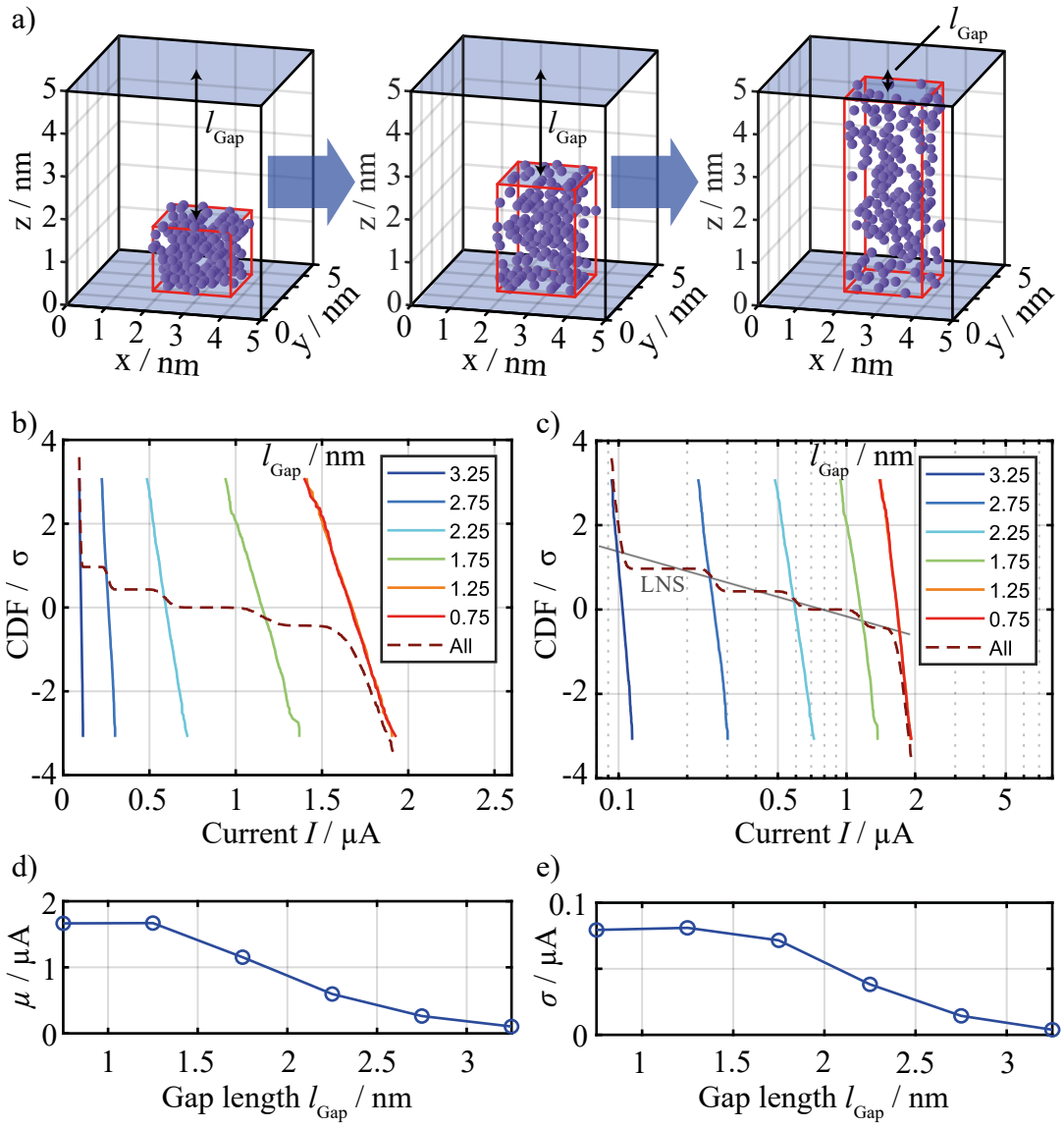

FIGURE 5.19: Static simulation for 200 defects with variation of the gap length l_{gap}. a) Exemplary configurations. b) Normally distributed read current distributions. A decreasing gap results in increased read current and tilting of the distribution towards higher σ. Note that σ here is the standard deviation of the normal distribution, accounting for linear current. The combined distribution for all gaps is given by the dashed line. c) data from b) on logarithmic current scale, revealing log-normal statistics (LNS) for the combined distribution, exhibiting discrete steps due to high number of defects resulting in few possible configurations. d) and e) Extracted trend of μ and σ. Both parameters increase with decreasing gap. At 1.25 nm the trend saturates to a plateau.

Partially reproduced with permission from S. Wiefels et al. [210], © 2020 IEEE.

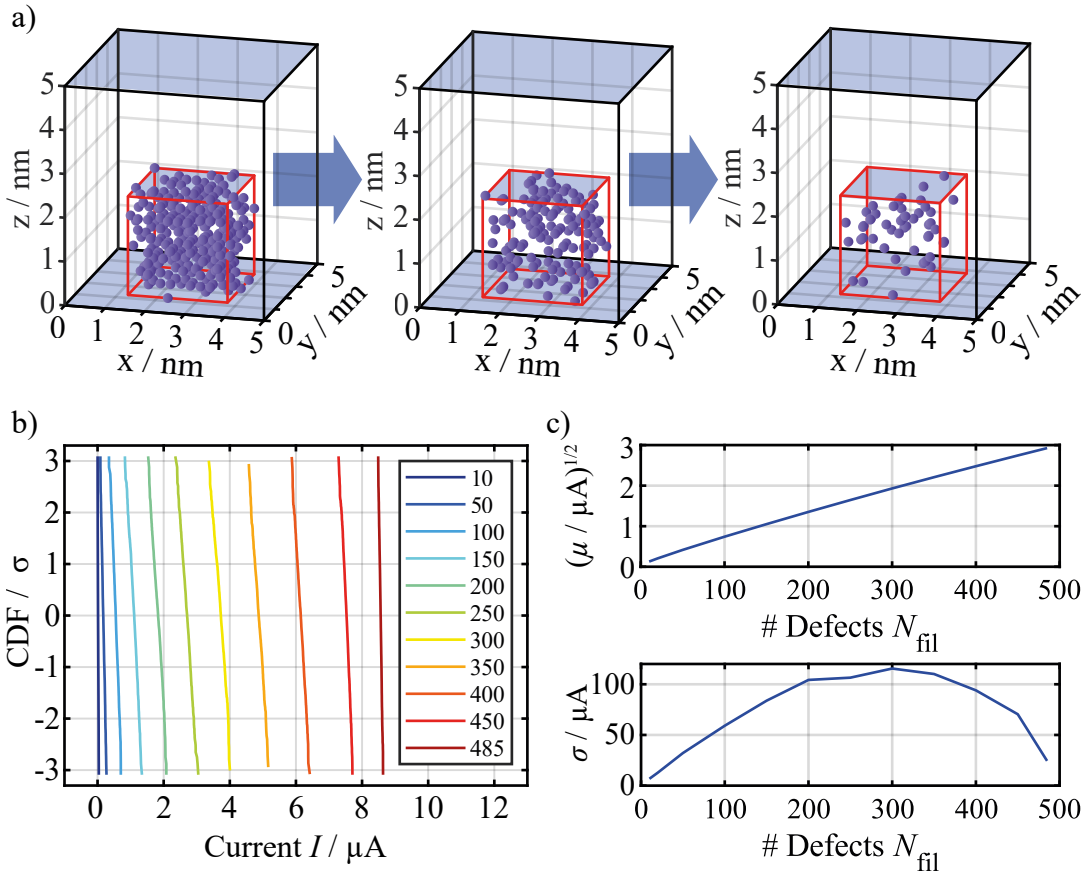

FIGURE 5.20: Static simulation for varying numbers of defects in a $(2.5 \cdot 2.5)$ nm^2 filament with a height of 2.25 nm. a) Exemplary configurations for 300, 150 and 50 defects. b) Normally distributed read current distributions. A decreasing defect concentration significantly decreases the read current. c) Extracted trend of μ and σ. μ increases quadratically with the number of defects. σ exhibits a maximum at 300 nm. Note that σ here is the standard deviation of the normal distribution, accounting for linear current.

experimentally observed degradation at high temperatures and long bake times.

The last aspect, investigated with the static version of the statistical model is a variation of the defect concentration. Figure 5.20, a) shows exemplary defect configurations with a number of defects decreasing from 300 to 50. The outer boundaries of the filament are kept constant. The resulting read current distributions are given in figure 5.20, b). As expected, the read current is strongly affected by the defect concentration. A decrease in the number of defects shifts the distributions to significantly lower current. As usual, the characteristic parameters μ and σ are extracted and displayed in figure 5.20, c). Regarding the current median, the square-root of μ is plotted which reveals a rather linear slope. Thus, μ increases quadratically with the number of defects in the filament volume.

The standard deviation σ is observed to have a maximum at 300 defects and from there decreases in both directions. In the statistical model σ is determined by the amount of possible (different) configurations of defects. With the given geometry, the

trend of σ indicates that the maximum number of configurations is apparent with 300 defects. If the number is increased, the decreasing number of available locations for defects lowers the number of possible configurations. Vice versa, a reduced number of defects results in less defects to contribute to the random configuration. The trade-off of these competing events determines the observed maximum.
It was discussed above that a decrease of the number of defects is considered as degradation mechanism rather than an increase. Regarding the experimental trends, the reduction of the defect concentration could explain the observed increase of σ, if the actual number of defects and the filament geometries are in the respective region. Though, the impact of N_{fil} on σ is comparatively low, considering the wide range of simulated concentrations. Nevertheless, μ is found to decrease significantly with decreasing N_{fil}. This contradicts the measured long term trend of the retention experiments but could be an explanation for the short term current decrease after programming. Here, a limiting factor could be the depletion of oxygen in the electrodes which is required for the recombination of oxygen vacancies. Alternatively, the number of oxygen vacancies close enough to the respective electrode to recombine could act as limiting factor.
In conclusion, the experimentally observed degradation (c.f. chapter 5.1.2) might be explained as follows: Initially, few oxygen vacancies recombine with oxygen of one of the electrodes until one of the mentioned limiting factors applies. As a result, the read current initially decreases. Subsequently, the diffusion of defects from the filament region towards the active electrode, resulting in a reduction of the gap, causes both long term degradation effects, i.e.

- monotonously increasing median μ, and
- tilting or broadeing of the distribution towards higher σ.

A reduced read current due to loss of oxygen vacancies is also considered in literature as typical origin of high temperature degradation [130]. Furthermore, increasing current due to diffusion of oxygen vacancies towards a depleted (gap) region has been reported [132]. This dissertation combines both approaches. Furthermore, the applied focus on statistics reveals a broadening of the distribution which is explained by the second process as well.
As discussed above, a broadening of the filament (increasing A_{fil}) seems to play no significant role for the retention of the tested devices. Besides the simulated trends of μ and σ for the variation of A_{fil} do not fit the experimental trends, it can be stated that the other two simulated processes (i.e. recombination of $V_O^{\cdot\cdot}$ and decrease of tunneling gap) are sufficient to explain the measured retention characteristics. Additionally, the simulated trends are collected in table 5.4. All in all, this dissertation emphasizes

TABLE 5.4: Conclusion of the simulated degradation effects (c.f. figure 5.15). For each observed degradation effect, the possible underlying processes are listed

$\mu \uparrow$	$\mu \downarrow$	$\sigma \uparrow$	$\sigma \downarrow$
(b)	(a), (c)	(b), (c)	(c)

to evaluate retention data regarding their statistics and therefore to analyze read current distributions for the characteristic parameters of their underlying statistics. The overview in table 5.4 should provide an orientation in the identification of the degradation processes underlying any observed retention characteristics.
As presented above, the model is able to reproduce the experimental results very well and provides a certain guidance towards finding the origin of the long term degradation. Though, the presented static version of the statistical model is limited to the manual variation of the boundary conditions for the generation of random defect configurations. In the following section, the model is extended to a dynamical version by inclusion of the KMC algorithm to confirm the degradation mechanisms predicted using the static model.

5.3.2 Dynamic Model

One of the greatest open questions regarding VCM ReRAM is the discrepancy between the high long-term stability (Retention) of the programmed state and the low short-term stability (Read Noise). Whereas ReRAM is justifiably considered non-volatile and exhibits room temperature stability up to 10 years, the state varies significantly on a millisecond scale as discussed in chapter 4. If both underlying processes are due to diffusion of oxygen vacancies (random walk) then it is to question why such different time scales are observed.
To answer this question and to confirm the findings of the static model, the latter is extended by including the KMC algorithm. The general concept is to investigate the dynamic diffusion of defects within regions of different activation energies. Figure 5.21 shows the schematic of the resulting dynamic model. In figure 5.21, a) the geometries are sketched. The cell used in the static model is now divided into three regions with distinct activation energies, i.e.

(i) a filament region (red) with very low activation energy of $E_A = 0.7$ eV,

(ii) a gap region (blue) with medium activation energy of $E_A = 1.0$ eV, and

(iii) the surrounding oxide with a high activation energy of $E_A = 1.5$ eV.

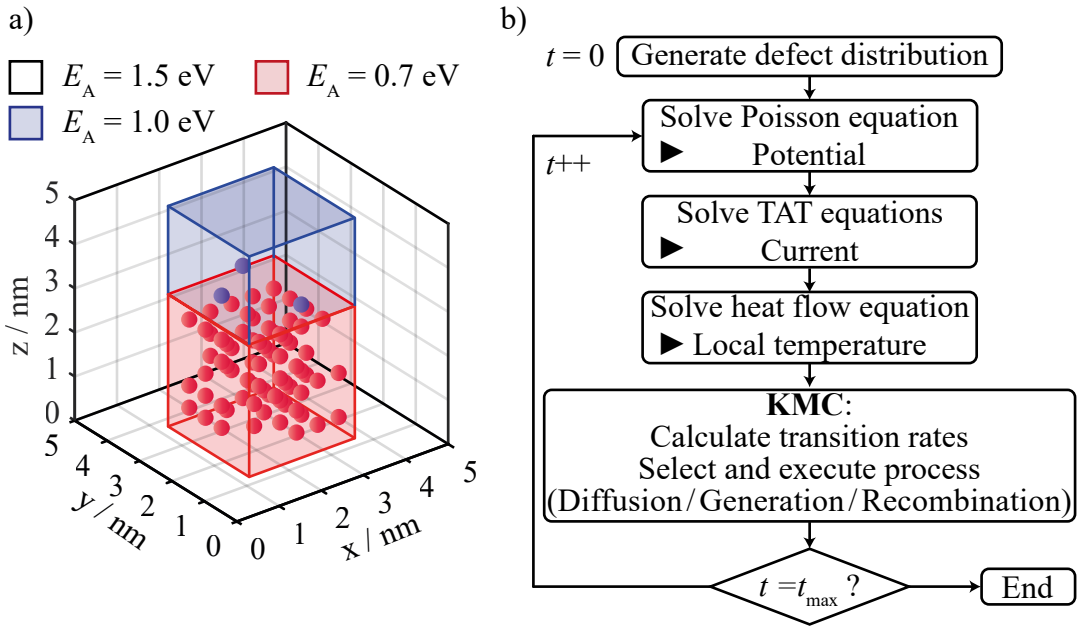

FIGURE 5.21: Schematic of the dynamic (KMC) version of the statistical model. a) Sketch of the underlying geometries. The cell is divided into 3 regions with different activation energies for defect diffusion. Within the filament (red), defects may diffuse easily with a low activation energy. Towards the gap region (blue) higher energy is required. The highest activation energy is implemented for the surrounding oxide (black). The cell is initialized with 50 defects placed randomly in the red filament region. Over time, some defects diffuse into the blue gap region. b) Flow Chart of the dynamic simulation.

Thus, defects may diffuse easily within the filament volume. At higher temperatures or longer bake times, the defects may also diffuse into region (ii), resulting effectively in a reduced tunneling gap. The least favorable region for defect diffusion is the surrounding oxide (iii) which should constrain the majority of defects to region (i) and (ii). The activation energy for the generation of oxygen vacancies is set to 2.8 eV. Respectively, the activation energy for their recombination is 1.5 eV. With the lower activation energies for the diffusion of defects within region (i) and (ii), these processes should be favored over generation and recombination.

The simulation flow is shown in figure 5.21, b). At $t = 0$ an initial defect distribution is generated regarding the generation rules used in the static model. Thus, 50 defects are placed on random sites within the filament volume (red, (i)). Analogously to the static model, the potential is calculated by solving the Poisson equation (2.5). The TAT equations (2.7)-(2.10) are solved to obtain the current for each defect. Additionally, the local temperature is calculated via the heat flow equation (2.15). Subsequently, the KMC module is used to determine the transition rates for all defects and possible processes (i.e. Diffusion, Generation or Recombination) based on the temperature of the system. According to the calculated transition rates, one process is executed. It may be noted that in this simulation no external voltage is applied to the cell to

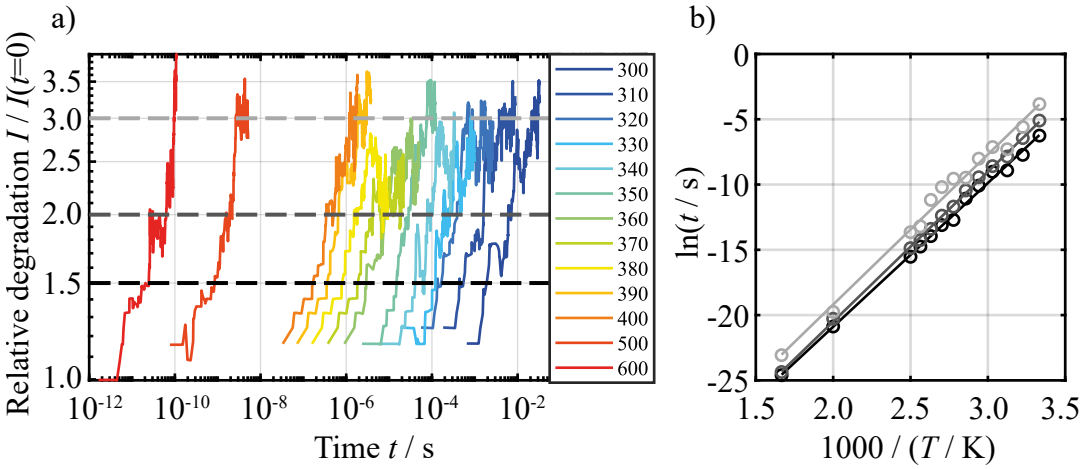

FIGURE 5.22: Results of the dynamic retention model. Simulations are performed at various temperatures from 300 K to 600 K a) Relative degradation (i.e. Current divided by initial current) reveals comparatively linear slopes on double logarithmic axises, superimposed by significant fluctuations due to random defect diffusion. The degradation time is extracted for 50 %, 100 % and 200 % current increase and added to the Arrhenius plot in b). The resulting activation energies range from 0.96 eV to 0.99 eV, corresponding to diffusion in region (ii).

simulate the degradation during a bake experiment analogously to the experimental data. Except the generation of the initial defect configuration, all steps are repeated until a defined simulation time t_{max} is reached. After each KMC step, the state of the cell is read out by applying 0.2 V with deactivated KMC module, equal to the static simulation. Thus, the state can be determined without having to account for an influence of the applied voltage to the retention.

The described simulation is executed for several temperatures ranging from 300 K to 600 K. The corresponding current traces are plotted on double logarithmic axes in figure 5.22. Additionally, the current is divided by the initial current to gain the relative degradation of the state. As expected, the temperature has an extensive impact on the time scale of the degradation. From 300 K to 600 K the time scale decreases by approx. 8 orders of magnitude. Besides, the traces are superimposed by comparatively large random fluctuations resulting from the random movement of defects in the system. Though, the slopes are rather linear on the double logarithmic scale and approx. parallel, suggesting a common activation energy.

To identify the mechanism which causes the degradation, the activation energy is determined from the degradation plots in figure 5.22, a). Therefore, the degradation time is extracted at 50 %, 100 % and 200 % current increase and added to the Arrhenius plot in figure 5.22, b). It can be seen that, despite the strong current fluctuations, the degradation times are in very good agreement with the Arrhenius fit. Here, the deviation from the fit line is most pronounced for the highest degradation of 200 %. This is reasonable with respect to the large current fluctuation in this range, observed

in figure 5.22, a).
The activation energies determined from the slopes of the Arrhenius fits are given in 5.22, b) and range from 0.96 eV to 0.99 eV. Since these values are very close to the programmed 1.0 eV in the gap region (ii), diffusion of defects into and in this region is identified as the limiting factor of the observed degradation effect. Whereas the diffusion of defects in the filament region, favored due to the low activation energy, may contribute to the superimposed current fluctuations, the degradation is clearly caused by defects which diffuse into the gap region and thus decrease the effective tunneling gap towards the active electrode. This is congruent with the findings obtained by the static model and therefore confirms the diffusion of defects from a limited filament region towards the active electrode as likely origin of the increasing current median μ during bake experiments. Additionally, the random current fluctuations in the dynamic model seem to increase with time which would result in a broader current distribution which could explain the tilting of the distributions towards higher standard deviation σ during bake experiments.
Finally, the dynamic model is used to study the coexistence of high long term stability (retention) and low short term stability (read variability, noise). Therefore, the simulation discussed above is repeated with 75 defects and for three different temperatures. The system is tested at a very low operating temperature of 250 K, room temperature (300 K) and a typical bake temperature of 500 K. The activation energies are equal to the ones used before (c.f. figure 5.21, a) for all regions but the blue gap region. It was found that a higher difference between the activation energies of filament and gap region is required to account for the mentioned discrepancy of the time scales for retention and noise. Therefore, the activation energy of the blue region (ii) is increased to 1.2 eV. The resulting current traces are plotted on the same time scale in figure 5.23, a). The plot reveals three significantly different behaviors: At the lowest temperature (250 K) the random current fluctuations are suppressed to large extend. This means, the diffusion of defects is frozen in and no ionic noise is observed. At room temperature the model produces current fluctuations between rather discrete levels as observed in typical noise experiments (c.f. section 4.2.1). Though, the current median seems to be stable on the presented time scale. This is in excellent agreement with experimental read noise. At the highest simulated temperature (500 K), the current fluctuations increase significantly and, most importantly, the current median drifts towards higher current, as observed during the bake in retention experiments. Thus, the presented dynamic model is able to explain consistently the coexistence of high retention time and room temperature instability.
Figure 5.23, b) shows two exemplary defect configurations corresponding to the orange current trace for 500 K in figure 5.23, a). At $t = 0$ µs all defects remain in

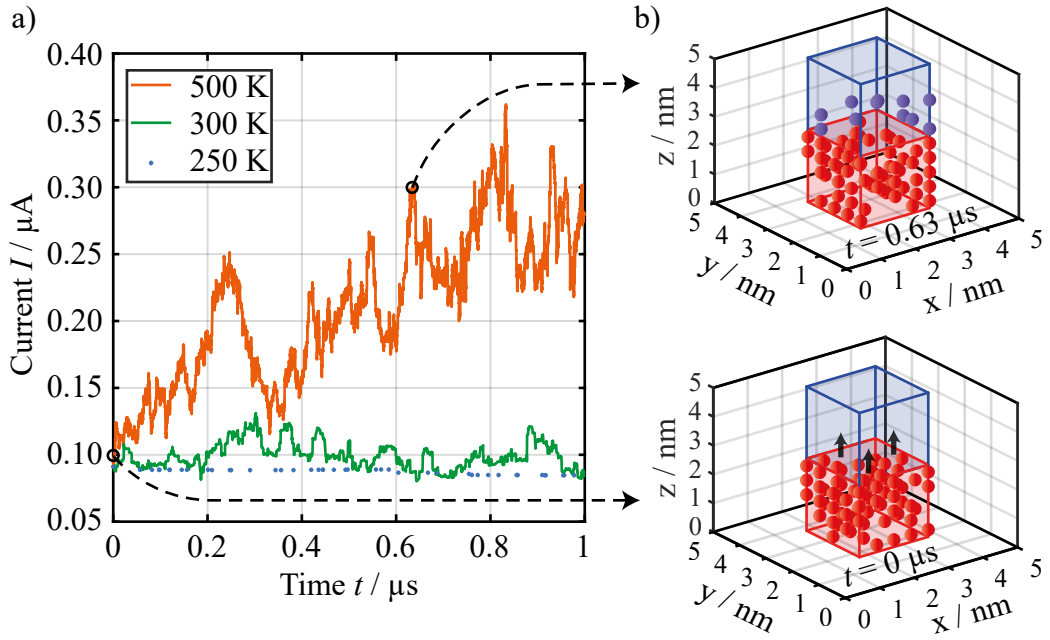

FIGURE 5.23: The dynamic model combines read variability and retention. a) Current traces simulated with 75 defects for low temperature (250 K), room temperature (300 K) and bake temperature (500 K). At 250 K current fluctuations are suppressed. At 300 K typical read noise with constant current median is observed. At 500 K the current median drifts additional to random current fluctuations. b) Exemplary defect configurations produced by the simulation for 0 µs and 0.63 µs at 500 K. It becomes clear that diffusion of defects into the blue gap region causes the current increase.

the confined, red filament volume. This results in the low initial read current. After 0.63 µs several defects moved into the blue gap region. This lowers the effective tunneling gap towards the active electrode and results in an increased read current, as observed for the orange current trace at $t = 0.63$ µs. This confirms the conclusion that a diffusion of defects towards the active electrode is the origin of the high temperature or long term degradation, limiting the retention time.

It may be remarked that the time scale in this simulation is comparatively small. Comparing figure 5.23, a) to the noise characteristics in figure 4.5, the frequency of the simulated room temperature noise (green) is approx. 8 orders of magnitude higher. The deviation of the simulation to realistic time scales is even larger regarding the orange, high temperature simulation. Retention times of several hours at 500 K (c.f. figure 5.4) are far above the time frame of 1 µs. However, at a given temperature, the time scale is directly determined by the programmed activation energies. Thus, the model could be adjusted to realistic times by increasing the activation energies of the respective regions. Yet, the determination of the real activation energies for all processes is challenging and requires comprehensive measurements of noise and degradation at several time scales and temperatures. The latter should cover high bake temperatures, room temperature and especially deep temperatures to determine

the activation energy of the room temperature instability. Simultaneously, the evaluated time scale should cover high frequency fluctuations (ns to s) as well as typical retention times of hours to days. Since, this data is not available in the frame of this dissertation, the activation energies of the dynamic model are not tuned toward most likely real values.
Another aspect is the temporal resolution which scales with the time scale. Regarding more realistic activation energies, the simulation could not resolve both room temperature instability and high retention. If the model was tuned to exhibit retention times of several hours at 500 K, the time increments would be far too large to resolve room temperature noise.
Hence, the parameters are chosen to present the general concept of multiple activation energies for the diffusion of defects in VCM cells. As presented, using regions characterized by different activation energies, the model is able to consistently cover both high retention and high read variability. With a comprehensive data set, the model could be used to estimate the activation energies of the underlying processes.
Despite the model providing consistent results, it should be discussed why those regions of different activation energies should occur in a VCM cell. Independently of this model, it is not yet fully understood why defects should remain in the vicinity of a conducting filament. The high concentration of defects in this region should provide a driving force for a radial diffusion of defects until the defects are distributed equally in the cell volume. Considering reasonable activation energies for the diffusion of oxygen vacancies in HfO_2, the actual retention should be significantly worse due to this effect. Nevertheless, the defects seem to stay in a confined filamentary region. Therefore, it seems reasonable to assume different activation energies for the diffusion inside the filament (red) and in or into the surrounding oxide (black). A possible explanation for this difference might derive from the electroforming process. The initial generation of the conductive filament is characterized by extremely high local current and temperature at the moment of dielectric breakdown [11]. It is conceivable that these conditions yield a local change of the crystal structure [211] resulting in different conditions for the diffusion of oxygen vacancies in the filament and the surrounding oxide.
However, this does not account for the difference between the red filament region and the blue gap region. A promising approach was published by Schie et al. [212]: Molecular dynamics simulations revealed sub-diffusive behavior in amorphous HfO_2. Using KMC methods this could be attributed to nanoscale confinement of the migrating defects. According to the study, it seems likely that the switching oxide is divided into several regions characterized by high ion mobility which are confined by higher diffusion barriers. The dynamic model discussed in this dissertation, represents a

strongly simplified implementation of the approach suggested by Schie et al. [212]. It is demonstrated that the concept is suited to model the coexistence of high retention and high room temperature instability. Further studies should refine the model by introduction of an extended number of high-mobility regions separated by higher barriers. In conclusion, this might be a promising approach to comprehensively understand both read variability and retention.

5.3.3 Consistency with Empirical Model

In the preceding sections, two different approaches to model the experimental retention characteristics are presented. On the one hand, the empirical model utilizes tunneling via a normally distributed gap (with limited minimal gap) to fit the measured read current distributions. On the other hand, the statistical TAT model relies on spatially resolved random configurations of defects. By adding a variability to the length of the gap l_{gap}, log-normal statistics are achieved. For comparison, figure 5.24 gives typical results of the statistical model in a) and the empirical model in b). As shown in figure 5.24, b) the results of the empirical model are in very good agreement with the experimental data. The model matches the log-normal, as well as the current limited regime. Figure 5.24, a) depicts read current distributions for different gap lengths l_{gap} simulated using the statistical TAT model. As discussed above, log-normal statistics are achieved by the combination these distributions, which results in a combined distribution with varying gap. At both ends of this distribution, determined by the borders of the simulated gap interval (i.e. 0.75 nm and 3.75 nm), the log-normal statistics are discontinued. Here, the combined distribution approaches the respective normal distribution of the corresponding gap (0.75 nm or 3.75 nm). This results in a similar current limited regime as observed experimentally. Though, this limitation is comparatively sharp due to the hard limit of the minimal or maximal gap length l_{gap}.
Although both modeling approaches utilize a variation of the gap length, the statistical model is by design limited to a spatial resolution of 0.5 nm. Therefore, the variation of l_{gap} is characterized by rather discrete steps. Despite the fact that the model is able to reproduce log-normal statistics, a finer variation of the gap would be beneficial for a comparison of empirical and statistical model. The limited resolution of the statistical model is well motivated by reduction of the simulation time and the desired limitation of the defect density (c.f. chapter 4.3.2). Though, filament / gap (or plug / disc) structures in a real, amorphous material are expected to be much less discrete. Consequently, the simulated data in figure 5.24, a) is extended by interpolation of the gap length to 0.01 nm increments. This is achieved by interpolating the

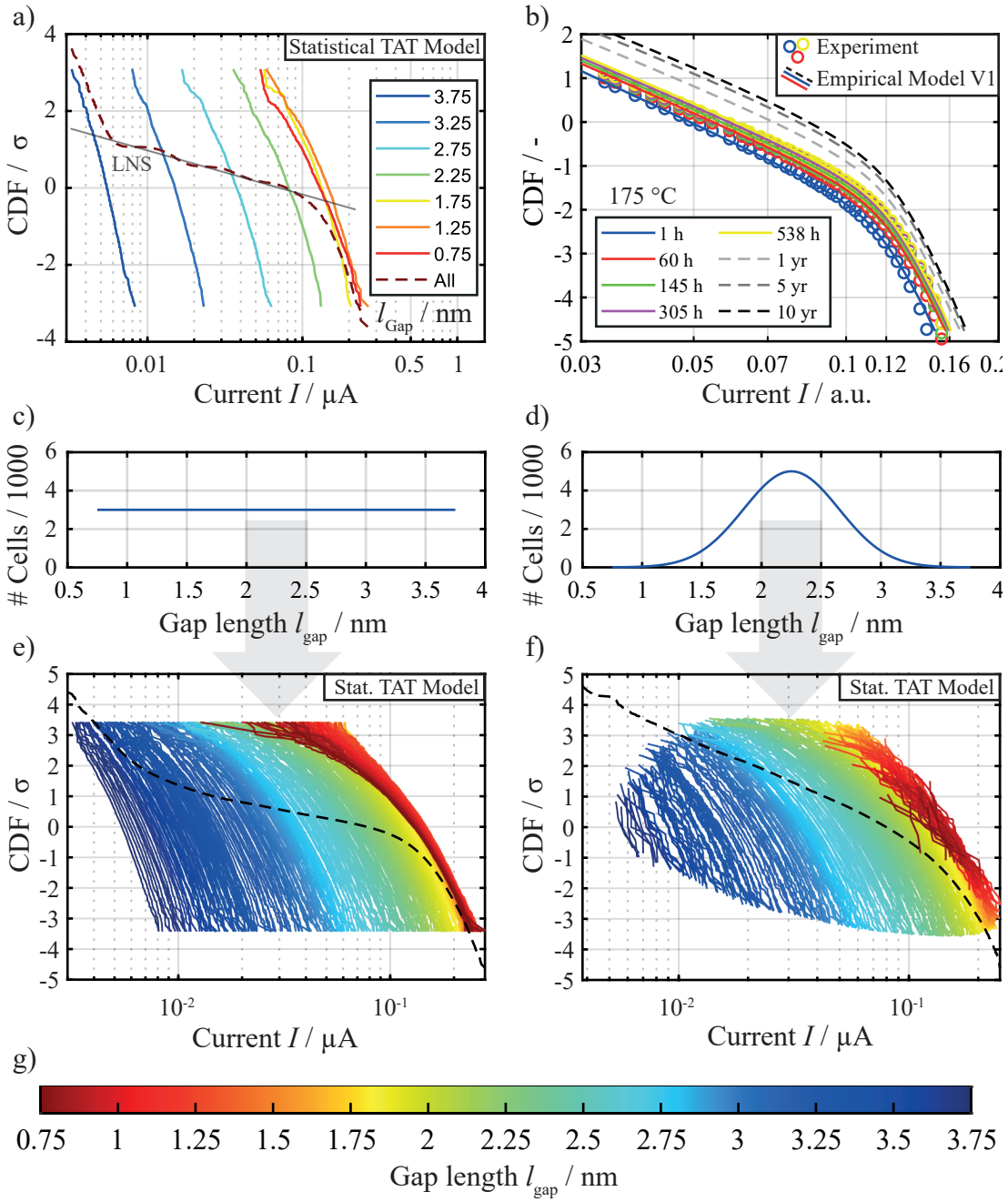

FIGURE 5.24: Comparison of empirical and statistical model. a) Exemplary results of the statistical model with variation of the gap length l_{gap}. The combined distribution (dashed) exhibits log-normal statistics and sharp limitations at both ends. b) Exemplary measured retention results for a bake at 175 °C, fitted by empirical model V1. c)-g) Extension of the statistical model by interpolating for finer increments of the gap length, with c) equally distributed l_{gap}, d) normally distributed l_{gap}. e) Read current distributions with equal gap distribution. The combined distribution (dashed) is smoothed compared to b), but still sharp limitations at the borders (0.75 nm to 3.75 nm) are observed. f) Analogous to e) with normally distributed gap. No bending at low current end is observed. A gradual current limit is observed for small gaps due to the reversing trend of μ for $l_{gap} < 1.25$ nm. The results show that a normally distributed gap explains the experimental observations convincingly and connects both model approaches.

extracted trends of μ and σ of the individual normal distributions (c.f. figure 5.18, d) and e)) and generating additional normal distributions for the desired l_{gap} with 0.01 nm increments.
The resulting distributions are shown in figure 5.24, e). The respective gap length can be identified by the color code in figure 5.24, g). Analogous to the originally simulated distributions in figure 5.24, a) the current median increases exponentially with reduction of l_{gap} until approx. 1.75 nm. In the red range of distributions ($l_{\mathrm{gap}} < 1.25$ nm) the trend is reversed. The reason for this might be either the higher mean distance between defects or the Fermi distribution of the electrode metal, as mentioned above. Additional to the finer increments of the gap length, the number of cells is increased to account for statistics in the range of the measured industrial cells.
Again, the linear distributions for different gap lengths are combined to one distribution depicted by the dashed, black line in figure 5.24, e). As expected, the curve is smoothed compared to figure 5.24, a). Though, the current limiting regimes at both ends of the distribution are still comparatively sharp.
The major difference here to the empirical model is that the gap lengths in the combined distribution are effectively drawn from a uniform distribution as depicted in figure 5.24, c). For each calculated normal distribution, 3000 cells are considered. Since the empirical model is based on normally distributed tunneling gaps, this concept is carried over to the statistical model. As shown in figure 5.24, d), normal distributions are generated for the same gap length with 0.01 nm increment as before, but the number of cells for each gap is adjusted to fit a normal distribution. Consequently, 5000 cells are considered for the median gap at 2.25 nm, with a decreasing number of cells in both directions of the gap length. The resulting read current distributions are shown in figure 5.24, f), with the corresponding color code in g). Here, the length of the individual lines corresponds with the number of cells included. Therefore, the field of normal distributions becomes thinner when the lower current limit is approached. On the opposite side this effect is not as clearly visible. This is due to the mentioned reversion of the trend for gaps below 1.25 nm. Thus, on the right side of the field, several lines lay on top of each other. Though, it can be seen that the dark red lines are shorter than the yellow section. As a consequence, the combined distribution (dashed black) is not symmetrical at both ends of the current spectrum anymore. With the right choice of a gap distribution, the bending of the distribution at low current can be suppressed. Thus, the sharp current limit due to the former discrete maximum of the gap interval is not visible. At high read currents, a bending is still observed which is now significantly more gradual, as observed experimentally. Here, the limitation due to the sharp lower gap limit is also prevented by the normally distributed gap lengths. Instead, the bending of the

combined distribution can be attributed to the inverting trend of the read current with decreasing gap.
This means that the former manually induced limits are replaced by a current limit originating from the physics of the statistical TAT model. It is remarkable that the symmetrical gap distribution results in an asymmetrical read current distribution.
All in all, this allows for two major conclusions:
Firstly, with the inclusion of normally distributed gaps, the results of the statistical model come even closer to the experimental observations. The model intrinsically explains the current limiting effect (bending) of the distribution at high read currents, without manually introducing borders.
Secondly, the general concept of a normally distributed gap is suited very well to account for the observed log-normal statistics. Even though the two presented models are substantially different, both are based on statistics and a variation of the gap length. If a normally distributed gap is also considered in the statistical model, the results of both models are largely consistent and in respectable agreement with the experimental data.

5.4 Conclusion

In this chapter, experimental retention results were presented for the HRS of two different VCM systems. Regarding ZrO_2, the high temperature degradation is characterized primarily by tilting (i.e. broadening) of the cumulative read current distribution. Compared to the academical ZrO_2 cells, the industrial HfO_2 based devices provide a significantly higher number of cells. This enabled a comprehensive analysis of the statistical aspects of the investigated retention. It was demonstrated that the degradation of the HRS is composed of tilting and shifting of the log-normal distributions. Therefore, the characteristic parameters μ and σ were evaluated quantitatively. Whereas μ initially decreases and subsequently increases with advancing bake time, σ increases monotonously and was observed to be the most critical factor regarding the retention of the read window. With broadening of the underlying distribution being the limiting factor of the retention time, it can be stated that the investigation of single cell retention is ineffective. Instead, the intrinsic statistics should be characterized.
The increased statistics of the industrial devices furthermore revealed a bending of the distribution at higher read current. This limiting effect could be attributed to the internal statistics. For given boundary conditions, there will be an optimal defect configuration, which results in the highest achievable read current.
Furthermore, two model approaches were presented: Firstly, an empirical model was

developed which enables to simulate read current distributions in excellent agreement with the experimental data. Additionally, it allows to extrapolate the measured retention characteristics toward operating temperatures and technically relevant retention times (e.g. 10 years at room temperature). Due to the low computational burden, these simulations can be extended to extremely low percentiles to determine the retention of the read window for very large memory arrays.
Secondly, a statistical model was presented which grants a deeper understanding of the processes underlying the observed degradation processes. Thus, the diffusion of defects from a confined filamentary region toward a tunneling gap at the active electrode could be identified as likely origin of both tilt and shift of the measured read current distributions during bake. The model also reproduced and explained the bending of the distribution at high current. Additionally, the dynamic version of the statistical model provided a first approach towards a model which consistently describes the long term stability (retention) and short term instability of VCM cells, despite their extremely different time (or temperature) scales.

6 Endurance

After variability and retention, the last key contributor to the reliability of ReRAM devices is the endurance or maximum number of possible SET / RESET cycles until device failure. This failure occurs either as RESET failure if the device becomes stuck in LRS or as SET failure if the device becomes stuck in HRS. Often the occurrence of one of these failures is caused by unbalanced programming conditions. If the SET process is too strong, the RESET may fail at some point or vice versa. It is therefore crucial to find the right programming parameters for the device under test. Based on the endurance scheme, introduced in chapter 3.3.3, an algorithm is developed in this work in order to maximize the endurance of the ZrO_2 based devices which will be presented in section 6.1.
However, even with ideal SET and RESET parameters, the endurance is observed to be limited and an irreversible (hard) endurance failure may occur. This hard failure does usually not appear as SET failure because deep HRS states can mostly be recovered by a stronger SET or even another forming operation. Instead, the end of life normally results from a very low LRS resistance causing a RESET failure. Stronger RESET conditions may then be futile, either because the voltage mainly drops over any series resistance or because the thermal stress due to high current becomes too high.
Guo et al. suggest that the final RESET failure is caused by continuous generation of excessive oxygen vacancies during cycling [7]. To what extend this happens, depends on the respective energy barrier, also called *defect formation energy* (DFE). Therefore, as explained in chapter 2.2.3, it is proposed to select material combinations with a high DFE in order to prevent RESET failures and thus ensure long endurance. The simulated findings of Guo et al. [7] are experimentally verified in section 6.2.
The results in this chapter are published in *Transactions on Electron Devices* [213].

6.1 Developed Endurance Algorithm

A first approach for an endurance algorithm for the ZrO_2 based devices investigated in this work was developed by Hüttemann [93]. The algorithm was based on the endurance scheme in figure 3.6. After every $n_{\text{cycles}} = 500$ cycles, the HRS and

LRS were evaluated. In case of endurance failure, different rescue operations were performed to recover operation of the device. This algorithm was observed to be very successful for devices which intrinsically showed good endurance characteristics. Moreover, the algorithm allowed a comparatively fast endurance assessment due to the high number (500) of cycles per pulse sequence [93]. However, this algorithm proved to be rather inappropriate for devices with weaker endurance characteristics. Thus, an improved algorithm was developed which is presented in the following.

6.1.1 Algorithm Details

In general, the concept of the endurance algorithm is to find the best programming conditions for each device. Furthermore, the endurance characteristics of each device are monitored over the applied switching cycles. Therefore, the SET and RESET parameters can be adjusted if the cycling characteristics change or the states drift towards too deep HRS or LRS. Here, a trade-off arises between time efficiency and the frequency of verify operations [160]. Whereas a high number of verify reads increases the chance of successful error correction, the measurement speed scales with the number of cycles which can be performed in pulse sequence. Thus, an algorithm is developed which dynamically adjusts the number of cycles per sequence to the success rate of the device under test, inspired by the work of Meng et al. [160]. Figure 6.1 provides an overview of the developed endurance algorithm. After electroforming, the SET voltage V_{SET}, the RESET voltage V_{RESET} and the number of cycles per pulse sequence n_{cycles} are initialized to $V_{\mathrm{SET}} = 0.5$ V, $V_{\mathrm{RESET}} = -0.5$ V and $n_{\mathrm{cycles}} = 0$. The initial voltages are kept intentionally low to allow the algorithm to find successful switching voltages with the lowest possible stress to the device. n_{cycles} will be increased once appropriate cycling parameters are determined.
As a first step, a SET with V_{SET} is applied and the resulting cell resistance R_{LRS} is read out and evaluated. If the resistance is above a threshold of 3 kΩ, V_{SET} is increased by 0.1 V and the SET is repeated. This loop is executed until the criterion $R_{\mathrm{LRS}} < 3$ kΩ is met. After a successful SET, the algorithm decreases the SET voltage again by 0.1 V if no increase of V_{SET} was necessary. This prevents the algorithm from continuously increasing the switching voltages over time and allows to find successful parameters with the lowest possible stress to the cell.
Then, the algorithm proceeds to the RESET process. Analogous to the SET operation, a RESET is performed and the resulting resistance R_{HRS} is compared to the target of 8 kΩ. Each time the RESET fails, the absolute of V_{RESET} is increased by 0.1 V. In case of successful RESET without voltage increase, $|V_{\mathrm{RESET}}|$ is decreased by 0.1 V. If the SET or RESET fails although V_{SET} or V_{RESET} is increased to a defined maximum

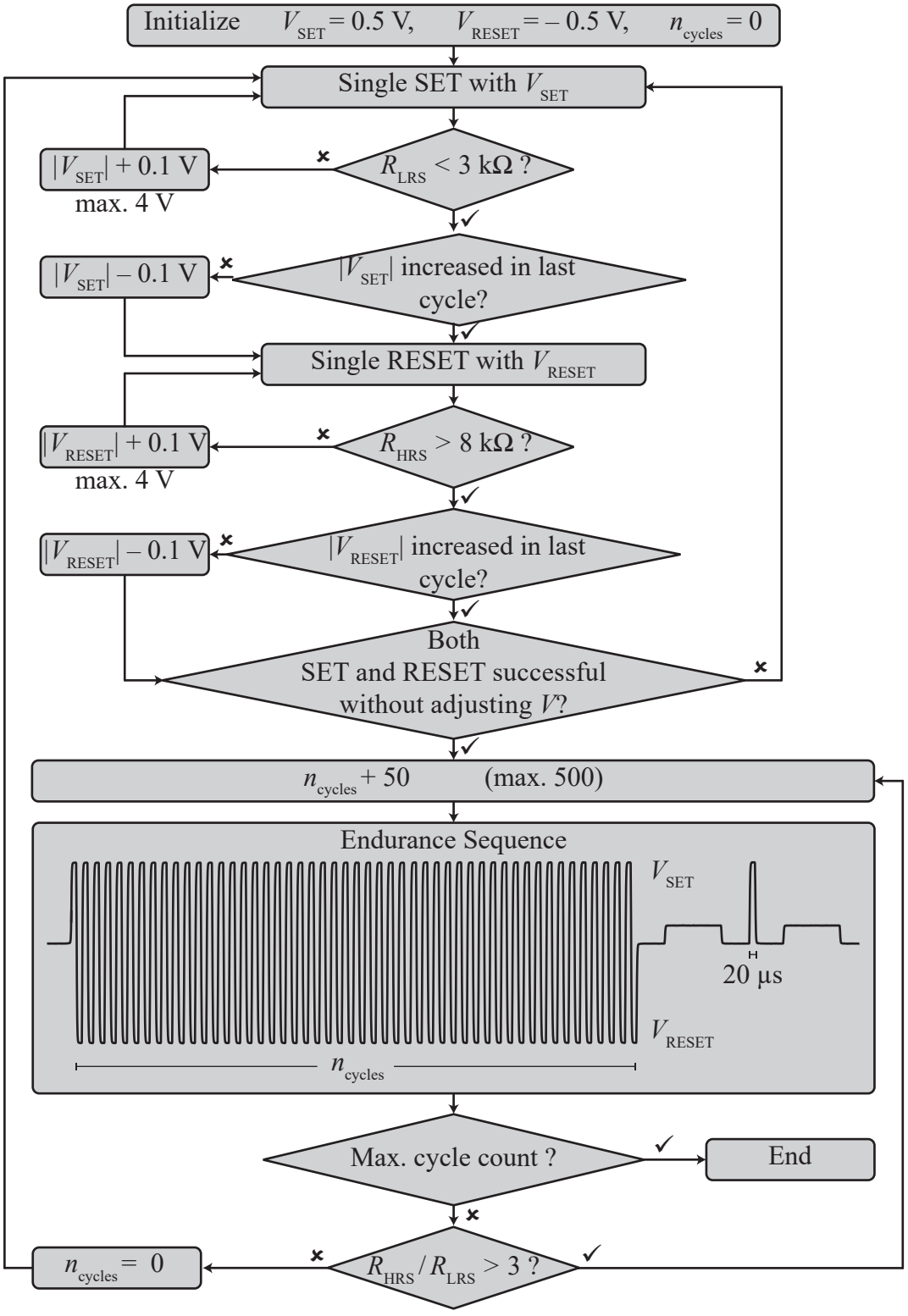

FIGURE 6.1: Flowchart of the developed endurance algorithm. In the upper part, appropriate switching parameters are determined for the tested cell. If SET or RESET fails, the respective voltage is increased. In case of successful switching, a voltage decrease is considered. In the lower part, the endurance scheme is executed. Here, the number of cycles per pulse sequence is increased each time the previous sequence was successful. This allows for frequent voltage adjustments if the cycling is unstable and fast endurance assessment in case of reliable switching. Adapted from [213].

of $\pm$ 4 V, the current number of cycles is saved as endurance and the algorithm is terminated.
After one successful cycle of SET and RESET, the algorithm checks if both processes succeeded without voltage increase. If not, the previous steps are repeated until appropriate parameters are established. Afterwards, n_{cycles} is increased by 50 (up to a maximum of 500) and the endurance scheme introduced in chapter 3.3 and figure 3.6 is executed. This means that n_{cycles} full cycles with the previously determined parameters V_{SET} and V_{RESET} are performed. Within the last cycle of this scheme, R_{LRS} and R_{HRS} are read out and evaluated. If the ratio $R_{\text{HRS}}/R_{\text{LRS}}$ exceeds the target of 3, the endurance operation is considered successful. In this case, the algorithm increases the number of cycles n_{cycles} per operation by 50 (up to a maximum of 500 due to the memory limitations of the tester) and repeats the endurance scheme. Additionally, a maximum number of cycles is defined at which the algorithm is terminated regardless of cycling success. It may be noted that the target read window of $R_{\text{HRS}}/R_{\text{LRS}} > 3$ is comparatively small. Aiming for a larger read window will presumably reduce the maximum endurance [164].
In the case that $R_{\text{HRS}}/R_{\text{LRS}} > 3$ becomes false, the algorithm resets n_{cycles} to 0 and jumps back to the single cycles with voltage adjustments. Thus, the algorithm is able to adjust the switching parameters if the endurance sequence fails. The frequency of these adjustments is tuned by n_{cycles}. The more stable the switching characteristics are, the more cycles per endurance sequence are executed. The algorithm therefore allows for both high measurement speed and frequent voltage adjustments if needed.

6.1.2 Results

The algorithm presented in the previous section is applied to the standard (30 nm Pt / 5 nm ZrO_2 / 20 nm Ta / 30 nm Pt) cross-bar devices investigated in this work (c.f. chapter 3.1). Exemplary results are depicted by figure 6.2. A statistical evaluation of the endurance is presented later in section 6.2. Figure 6.2 shows the endurance of three individual cells. Each plot shows the measured HRS and LRS resistances versus cycle count in the upper panel. The graph below shows the applied switching voltages V_{SET} and V_{RESET} at the respective cycle.
In general, it shows that a cell endurance of 10^6 to 10^7 can be achieved in the investigated material system using the presented algorithm. In more detail, it can be stated that the endurance characteristics are comprised by very stable regions and more unstable regions where the algorithm has to adjust the switching parameters frequently to prevent early endurance failures. In figure 6.2, a) the switching characteristics appear slightly unstable in the beginning. Using static programming

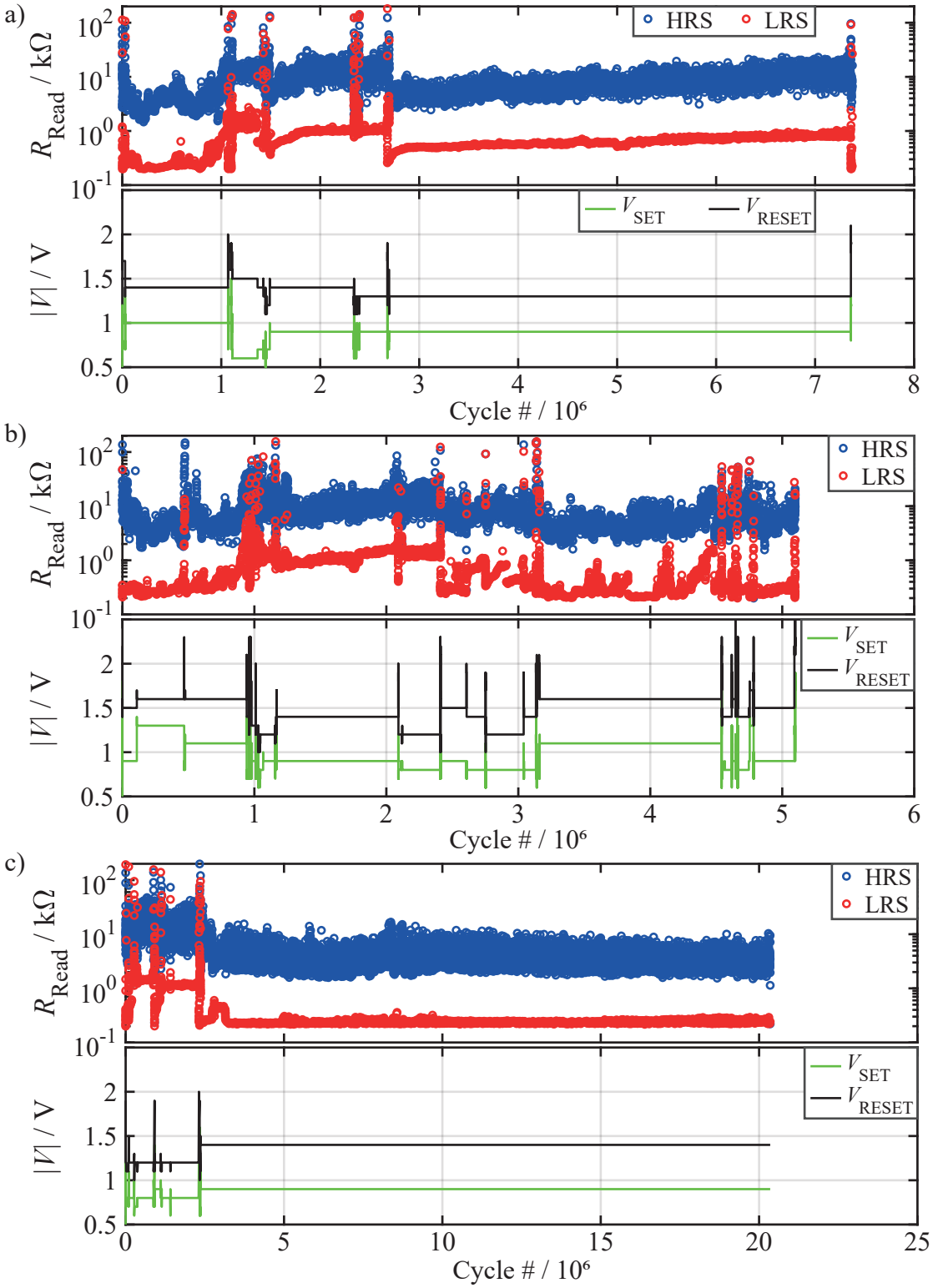

FIGURE 6.2: Exemplary endurance results for three cells of the standard (30 nm Pt/5 nm ZrO_2/20 nm Ta/30 nm Pt) cross-bar structure. In each part, the upper plot shows HRS and LRS resistance versus the cycle number. Below, the voltage applied by the algorithm at the respective cycle is shown. Irreversible RESET failure occurs after a) $7.3 \cdot 10^6$ cycles and b) $5.1 \cdot 10^6$ cycles. No terminal failure is observed in c). The measurement is terminated at the maximum cycle number. Partially reproduced with permission from S. Wiefels et al. [213], © 2021 IEEE.

conditions, endurance failure would have occurred several times within the first few hundred cycles and in the range of one to three million cycles. Thus, the algorithm prolonged the life of the tested cell significantly. After approx. $3 \cdot 10^6$ cycles, the cycling characteristics become extremely stable. Here, V_{SET} and V_{RESET} can be kept constant and n_{cycles} is automatically increased to its maximum of 500 cycles per pulse sequence. At approx. $7.3 \cdot 10^6$ cycles the cell becomes stuck in a comparably low ohmic LRS state and the RESET fails despite increased V_{RESET}. This denotes the final endurance failure and end of life of the tested cell.
In contrast to the cell in figure 6.2, a), a second cell shown in figure figure 6.2, b) never reached a sustained stable state. Instead, R_{HRS} and R_{LRS} fluctuate significantly from cycle to cycle causing several reversible (soft) endurance failures. Nevertheless, the algorithm managed to keep this rather unstable cell alive for up to $5 \cdot 10^6$ cycles. Again, a final RESET failure left the cell stuck in LRS and terminated the endurance.
Figure 6.2, c) shows one of the best cells, tested in this work, with respect to endurance. Here, the characteristics are notably similar to the first cell in figure 6.2, a). Within the first $3 \cdot 10^6$ cycles, several soft endurance failures occur which have to be corrected by the algorithm. Afterwards, R_{HRS} and R_{LRS} accumulate in very stable bands resulting in a cell endurance of more than $2 \cdot 10^7$ cycles. Finally, the defined maximum cycle number was reached and the measurement terminated although the cell was still operating successfully.
All in all, the presented endurance measurements show very typical characteristics for the investigated ZrO_2/Ta material system. In most cases the cycling is initially rather unstable and early endurance failures are likely. Therefore, an algorithm as the one presented in this work is especially useful to correct early failures and to find switching parameters which enable sustained stable switching.
The combination of initially unstable switching followed by a more reliable regime is also reported for industrial ReRAM devices [214]. Panasonic developed different endurance schemes ranging from a *relaxation technique* [215] without verify, over a conventional program-verify approach [216] to a hard verify with significant distribution shaping [214]. Although the latter causes significant stress to the cycled device, the best endurance is achieved if this hard verify is applied during the first 1000 cycles followed by the *relaxation technique* [215] without any verify.
The algorithm in this work is initiated with low voltages and tries to find the lowest possible biasing conditions which allow successful switching. This contradicts the approach with initially 1000 hard verify cycles by Panasonic [214]. In the further development of the endurance algorithm, stronger SET and RESET pulses within the first cycles should therefore be taken into account.

6.2 Influence of the Ohmic Electrode

In addition to the presented endurance results for the standard (30 nm Pt / 5 nm ZrO_2 / 20 nm Ta / 30 nm Pt) cells, devices with exchanged ohmic electrode are tested. According to the simulations by Guo et al. [7], the likely-hood of irreversible RESET failures is determined by the ohmic electrode metal as discussed in chapter 2.2.3. The RESET failure is believed to be caused by excessive generation of oxygen vacancies during cycling [7]. Thus, a high energy barrier or defect formation energy (DFE) is desirable for high endurance. It is therefore expected that ohmic electrode metals with lower oxygen chemical potential result in a higher DFE and hence higher endurance [7].
To verify the theoretical findings of Guo et al. [7], the endurance experiments are extended to cells with Hf, Zr or Ti replacing Ta as ohmic electrode metal. The expected trend of the DFE and thus endurance is Hf < Zr < Ti < Ta.
In order to obtain comparable endurance data, two different approaches can be considered. Either the programming conditions are equal for all cells or the parameters have to be optimized for each material system separately. The straight forward approach with equal parameters seems to produce the most comparable results. However, this would mean that V_{SET} and V_{RESET} are optimized for one stack. A worse endurance for the other stacks could therefore be due to imbalanced programming conditions. Hence, it becomes impossible to determine which material system is ideal with respect to endurance.
Thus, the second approach is used in this work. By determining the ideal programming conditions for each stack, the best possible endurance for each material system is obtained. Thus, the maximum endurance for each stack can be compared to find the best material system. Here, the developed algorithm plays a significant role since it is designed to find the optimal programming parameters for each tested cell.
Prior to the statistical evaluation of the endurance, exemplary endurance results for the devices with Hf, Zr or Ti as ohmic electrode metals are presented in figure 6.3. Analogous to the previous results for the standard ZrO_2/Ta devices, the HRS and LRS resistances are shown in the upper panel and the respective SET and RESET voltages in the plot below. Figure 6.3, a) shows typical endurance characteristics for a cell with Hf as ohmic electrode metal. As expected according to Guo et al. [7] the maximum number of cycles is significantly lower compared to the Ta devices. Several soft endurance failures occurred which could be corrected by the algorithm. After approx. 38,000 cycles the endurance is terminated by a very low resistive LRS state resulting in permanent RESET failure.
Similar results are obtained for cells with Zr as ohmic electrode metal as exemplarily

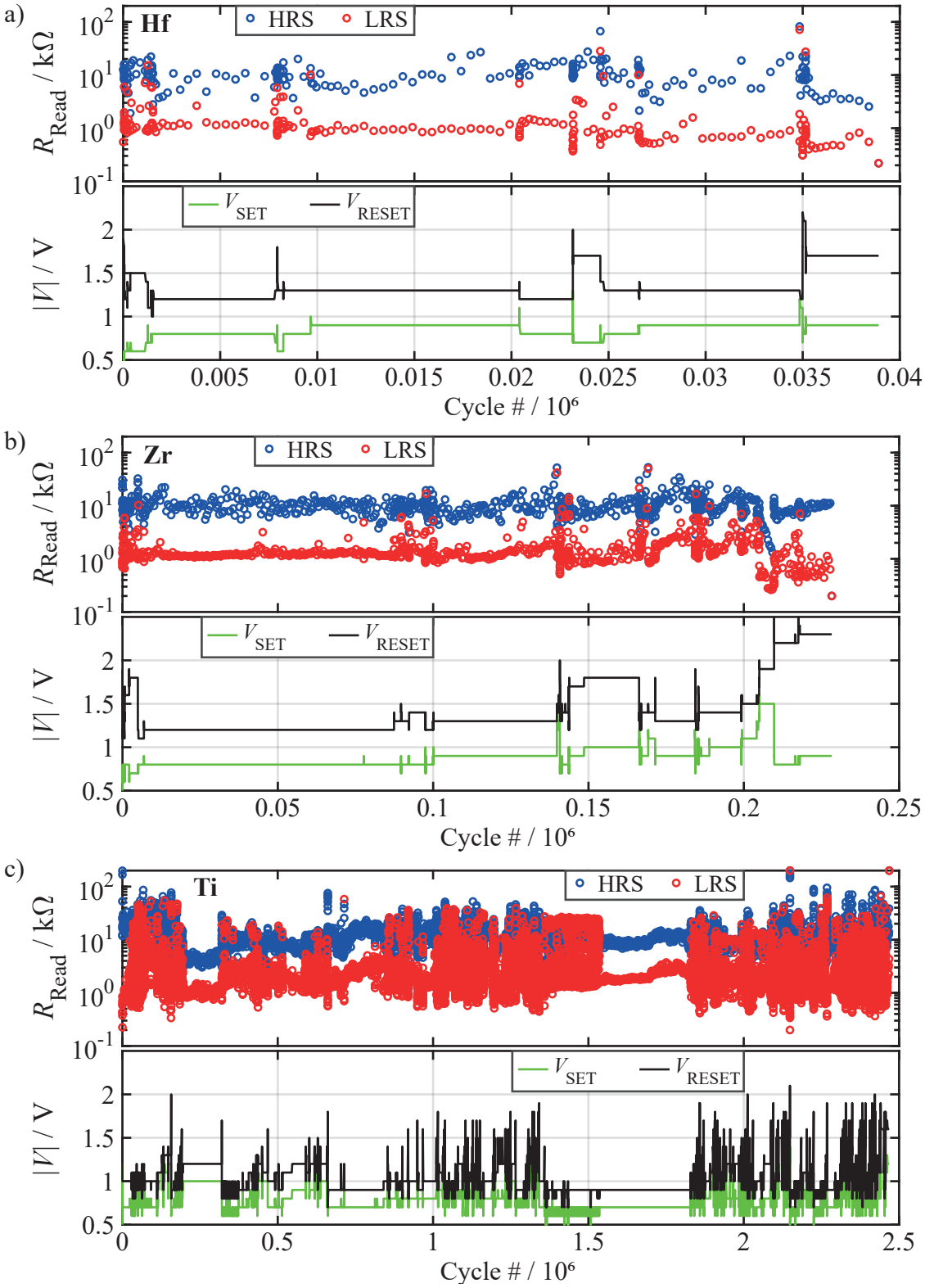

FIGURE 6.3: Exemplary endurance results for cells of the stack (30 nm Pt / 5 nm ZrO_2 / 20 nm *X* / 30 nm Pt), with *X* being a) Hf, b) Zr and c) Ti. In each part, the upper plot shows HRS and LRS resistance versus the cycle number. Below, the voltage applied by the algorithm at the respective cycle is shown. Irreversible RESET failure occurs after a) $3.8 \cdot 10^4$ cycles, b) $2.3 \cdot 10^5$ cycles and c) $2.5 \cdot 10^6$ cycles. Reproduced with permission from S. Wiefels et al. [213], © 2021 IEEE.

shown in figure 6.3, b). After multiple corrected soft endurance failures, a hard RESET failure terminates the measurement. However, the endurance is increased by nearly one order of magnitude compared to the ZrO_2/Hf device which is also in line with the simulation of Guo et al. [7].
Finally, an exemplary endurance trace for a device with Ti electrode is given in figure 6.3, c). Very unstable switching characteristics are observed for the ZrO_2/Ti devices and therefore a high number of voltage adjustments is applied by the algorithm. Nevertheless, the endurance is increased by another order of magnitude with respect to the ZrO_2/Zr device. Only the ZrO_2/Ta devices endured more cycles until hard failure. Thus, the endurance of the presented examples follows the expected trend Hf < Zr < Ti < Ta.
However, as stated throughout this work, reliability aspects should not be investigated on single cells. Instead, a statistical evaluation is required to reliably verify the theoretical findings of Guo et al. [7]. Therefore, the study is extended to 50 cells of each material stack. Here, the experimental setup with 32 contacted cells plays a significant role. With the measurement routines outlined in chapter 3.3, all contacted cells can be formed and tested automatically. Afterwards, the endurance algorithm is applied to all functioning cells. By this, the endurance of several cells can be determined automatically in a reasonable time and with low work effort.
As a result, cumulative distributions of the endurance are given for the four tested material systems in figure 6.4, a). Although the single distributions are comparatively wide, covering two to three orders of magnitude in endurance, it is observed that the general trend and thus the median follows the expected trend Hf < Zr < Ti < Ta. Since the difference in calculated DFE for ZrO_2/Hf and ZrO_2/Zr is comparatively small, it is reasonable that both endurance distributions are similar and close to each other. Nevertheless, the endurance is slightly higher with Zr as ohmic electrode metal, following the prediction of Guo et al. [7]. As expected, the highest endurance is observed with the standard ZrO_2/Ta stack. Here, the distribution is cut at the upper edge of 20 million cycles which equals the maximum number of cycles defined for the endurance algorithm. Probably, higher maximum endurance values could have been achieved here.
Additional to the general trend of defect formation energies (Hf < Zr < Ti < Ta), the respective numbers are extracted from [7] and summarized in table 6.1. It may be

TABLE 6.1: Defect formation energies extracted from [7]

Material	Hf	Zr	Ti	Ta
DFE / eV	0.1	0.2	1	1.65

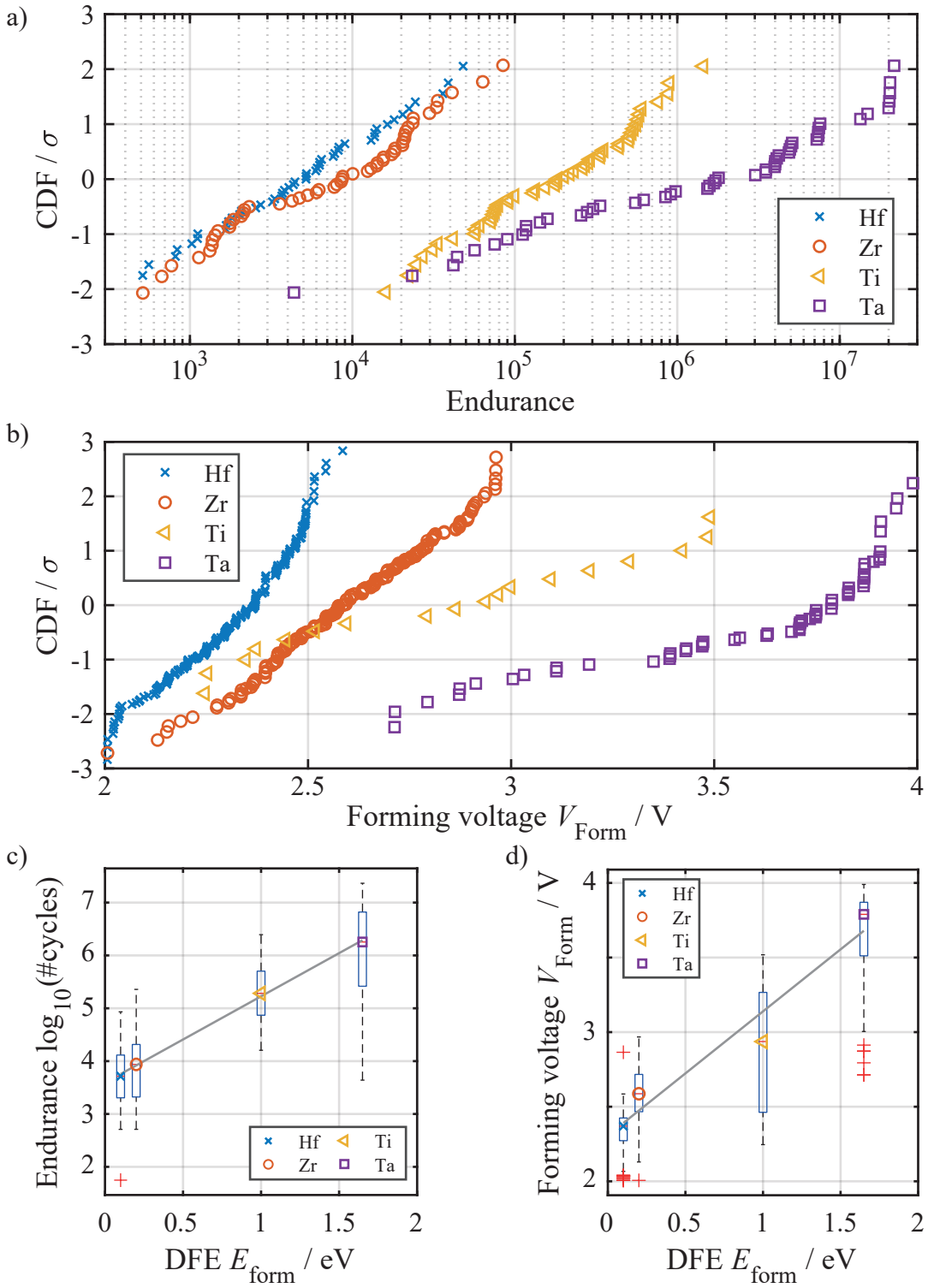

FIGURE 6.4: a) Cumulative probability of the endurance for Hf, Zr, Ti or Ta as ohmic electrode metals. b) Forming statistic for the same material stacks. c) The logarithm of the endurance linearly follows the DFE. d) The forming voltage linearly increases with DFE. A trade off arises for materials selection: Ohmic electrode metals with lower oxygen chemical potential result in higher DFE which increases both endurance and forming voltage. Data in b) taken from [93]. DFE in c) and d) extracted from [7]. Reproduced with permission from S. Wiefels et al. [213], © 2021 IEEE.

noted that the numbers are calculated for each metal with respect to HfO_2 with a DFE in the bulk oxide of $E_{\text{form,bulk}} = 5.9$ eV. However, since HfO_2 and ZrO_2 are nearly identical with respect to their physico-chemical properties, the difference is expected to be minor. For the generation of oxygen vacancies in bulk ZrO_2 values of $E_{\text{form,bulk}}$ in a range of 5.4 to 6.4 eV are reported [217]. Foster et al. determined $E_{\text{form,bulk}} = 5.94$ eV for m-ZrO_2 [218, 219]. Nevertheless, the relative trend for the four metals should not be affected by an offset in $E_{\text{form,bulk}}$.

Accordingly, the median endurance is extracted from the distributions in figure 6.4, a) and plotted versus the DFE in figure 6.4, c). Interestingly, the logarithm of the maximum cycle number not only reproduces the general trend of the DFE, but follows a nearly perfect linear dependency. It is therefore concluded that the measured endurance is in excellent agreement with the simulated data by Guo et al. [7]. It is experimentally verified that irreversible RESET failures as the key challenge in the device endurance can be significantly reduced by the choice of ohmic electrode metals with a low oxygen chemical potential. In this dissertation, the median of the endurance is increased by three orders of magnitude by Ta as ohmic electrode metal instead of Hf.

Potentially, the endurance could be further increased by exchanging Ta for a metal with even lower oxygen chemical potential as for example W or Ru which could be subject to future investigations. However, as a higher DFE energy hinders the generation of oxygen vacancies, it should also affect the forming voltage. Thus, forming statistics were recorded for the four discussed stacks by Hüttemann [93]. The resulting distributions are plotted in figure 6.4, b). As expected, the forming voltage is observed to increase with DFE. Unfortunately, only few of the devices with Ti as ohmic electrode were operational in this work. Nevertheless, the respective distribution fits to the expected trend. Accordingly, the median of the forming voltage is plotted versus the DFE in figure 6.4, d). Here, the forming voltage seems to follow a linear trend with the DFE. The deviation of Ti from the linear trend could be explained by the limited statistics due to the low number of functioning cells. These results confirm the conclusion of hindered oxygen vacancy generation by increased DFE.

Although a further increase of the DFE could lead to even higher endurance, the forming voltage would increase as well which can become an issue if the applied current compliance during forming is not ideal. Here, a higher forming voltage may cause higher parasitic currents in the moment of dielectric breakdown which could cause irreparable damage to the cell. All in all, the positive effect of an increased DFE on the endurance comes along an increase in forming voltage. This trade off has to be taken into account for material selection. Kim et al. investigated Ta_2O_5 devices and

reported a similar trend regarding the DFE. With W as ohmic electrode an improved endurance along with higher forming voltage was reported [168]. Furthermore, the higher DFE was observed to result in a larger read window [220].
Apart from the forming voltage, it seems likely that the programming voltages (V_{SET} and V_{RESET}) required for successful switching are affected by the DFE. As the programming algorithm adjusts both parameters to the cycling characteristics of each cell, a trend may be detectable regarding the DFE. Therefore, cumulative distributions of all occurring SET and RESET voltages are given in figure 6.5, a) and b). Compared to endurance and forming voltage, the impact of the DFE on the required switching voltages is significantly lower. This indicates that the resistive switching mainly occurs at the active (Pt) electrode. Respectively, the impact of the ohmic electrode is limited.
However, the devices with Zr or Hf as ohmic electrode and a rather low DFE are observed to require slightly higher RESET and lower SET voltages compared to Ti and Ta. Respectively, the ratio $V_{\mathrm{SET}}/V_{\mathrm{RESET}}$ is slightly decreased with lower DFE, as demonstrated in figure 6.5, c). Thus, the algorithm automatically shifts the bias towards the RESET direction to counteract the parasitic generation of excessive oxygen vacancies, if necessary. On the one hand, this underlines the effectiveness of the presented algorithm. On the other hand, this trend is in very good agreement with the simulation results of Guo et al. [7].

6.3 Conclusion

In this chapter, a sophisticated endurance algorithm is presented which allows for a fast and automated assessment of endurance. On the one hand, the algorithm is able to correct early endurance failures and aims to find the ideal switching parameters for each cell. On the other hand, the adaptive number of cycles per pulse sequence addresses the trade off between measurement time and frequency of voltage adjustments. The measurement speed is increased if a cell switches reliably and decreased if frequent corrections are required.
The algorithm is initiated with low voltages and tries to find the lowest possible voltages for successful switching to limit the stress to the device. Although this was demonstrated to be a successful approach to enhance the device endurance, the results by Panasonic suggest a potential improvement. Hard verify cycles within the first 1000 cells and subsequently softer programming may further improve the presented algorithm which should be aspect of further studies [214].
The second part of this chapter addresses the influence of the ohmic electrode with respect to endurance. Here, the theoretical results by Guo et al. [7] are experimentally

verified. It is demonstrated that ohmic electrode metals with lower oxygen chemical potential lead to a higher DFE and thus hinder the generation of oxygen vacancies. This results in less terminal RESET failures and therefore prolonged endurance. However, the hindered generation of oxygen vacancies also increases the forming voltage. This trade off has to be taken into account for the materials selections.

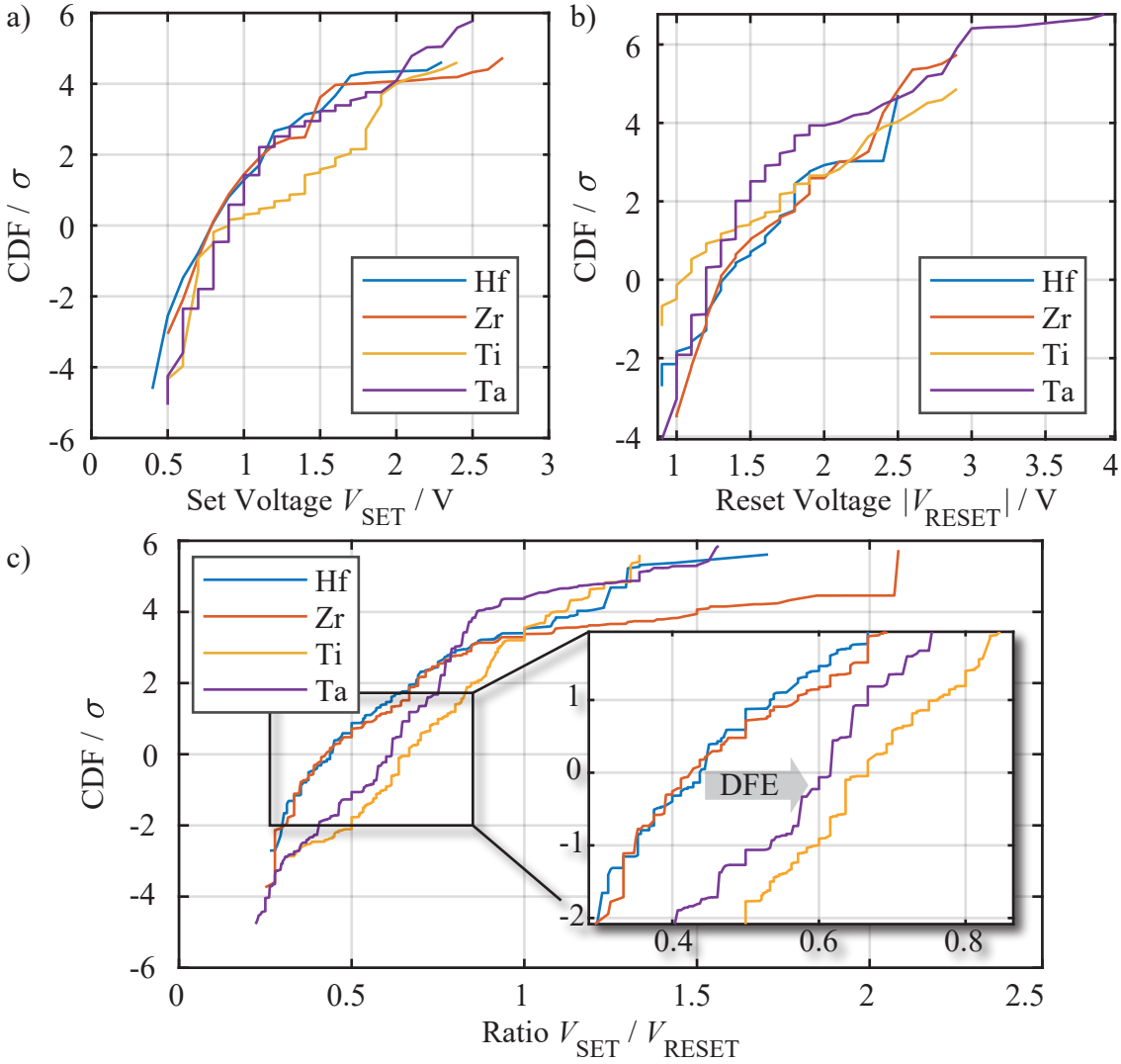

FIGURE 6.5: CDFs of the a) SET and b) RESET voltages determined by the programming algorithm. c) Ratio V_{SET}/V_{RESET}. The devices with Hf and Zr require slightly higher RESET and lower SET voltages to counteract excessive oxygen vacancy generation due to the low DFE, as highlighted in the inset. Reproduced with permission from S. Wiefels et al. [213], © 2020 IEEE.

7 Conclusion

The subject of the presented work was the reliability of VCM type ReRAM devices. Here, the focus was to proceed from the characterization of single devices towards large sets or arrays of cells. By this, statistically sound results were obtained but, more importantly, the intrinsic statistics of the tested devices could be characterized. This especially allowed, for the first time, to link reliability aspects to changes in the statistical parameters.
This dissertation covered two material systems, one based on ZrO_2, the other based on HfO_2 as switching oxide. On the one hand, a novel measurement setup was developed for the statistically sound characterization of the self-fabricated ZrO_2. It allows to probe and characterize 32x1 line arrays with a high level of automation due to the developed programming environment. On the other hand, experimental data of up to 2.5 million industrial HfO_2 was investigated in this work which provides exceptional opportunities for a university environment. Due to the physico-chemical similarity of the two oxides, comparable results were obtained to a large extend. However, differences between the industrial HfO_2 and ZrO_2 which was fabricated under laboratory conditions were pointed out.
With regard to the reliability of the investigated systems, three major aspects were covered, being their variability, the data retention and the cycling endurance. The respective results are summarized in the following.

7.1 Variability

The variability or stochastic nature of VCM ReRAM remains one of the largest challenges for their large scale adaption in industrial applications. In chapter 4, three different types of variability were presented, being device to device (D2D), cycle to cycle (C2C) and read to read (R2R). The variability between single devices (D2D) seems to be dominated by the stochasticity of the switching process. Large fluctuations in the resulting cell resistance occur from cycle to cycle and thus the C2C variability becomes indistinguishable from D2D variability. Since these types of variability can be significantly reduced by smart programming algorithms, the short term instability of the once programmed states (R2R) remains as key challenge.

It was demonstrated in this work that R2R variability determines the intrinsic statistics of a large set of cells and effectively limits the read window between LRS and HRS. Whereas the metallic nature of the conduction in LRS results in a normal distribution of read currents or resistances, the HRS usually is log-normally distributed. Via an empirical model as well as KMC methods, the most likely origin of these statistics was found to be tunneling across a normally distributed gap in HRS. Here, the exponential dependence of the read current on the tunneling gap or barrier results in the observed log-normal statistics.

7.2 Retention

These findings could directly be transported towards the understanding of long-term degradations which determine the retention of a memory array. Since the read current of individual devices fluctuates already on very short time scales (ns - µs), the retention can not be reliably characterized by tracing single devices. Instead, it was demonstrated that the long-term degradation appears as change of the intrinsic statistics or R2R variability. Firstly, the whole distributions were observed to shift over time either to lower or higher read current. The former is explained by loss of oxygen vacancies due to recombination, the latter by diffusion of oxygen vacancies from a plug or filament region towards a gap or disc region.
Secondly, the width of the distribution was observed to increase over time. This broadening or increase in standard deviation occurs in both LRS and HRS and is identified as major challenge regarding retention. Conclusively, the remaining read window is mainly limited by this process. Here, two different models were developed. On the one hand, an empirical model assuming tunneling across a normally distributed gap fits the experimental data very well. It represents a novel approach for the extrapolation of high temperature measurements towards the required retention time at lower operating temperatures. On the other hand, a statistical model was developed. It reproduces the intrinsic variability of the investigated devices by random configurations of oxygen vacancies within defined boundaries. It was demonstrated for the first time that a change in these boundaries can explain the observed degradation effects. A loss of oxygen vacancies leads to the expected current decrease. An expansion of the filament boundaries towards the active electrode accounts for both current increase and broadening of the distribution.

7.3 Endurance

In order to improve the endurance of VCM ReRAM two approaches were presented in this work. Firstly, a novel programming algorithm was developed which combines fast assessment of the cell endurance with frequent adjustments of the programming parameters. In order to perform 10^6 and more cycles within reasonable measurement time, as many cycles as possible are applied within one pulse sequence. However, the number of cycles applied at once needs to be reduced if the switching characteristics are unstable. The algorithm dynamically adjusts the SET and RESET voltage, as well as the number of cycles per pulse sequence in order to achieve the maximum endurance within reasonable experiment time.
Secondly, this algorithm is used to determine the maximum endurance for cells with varying ohmic electrode metal. Guo et al. [7] predicted from ab-initio methods that ohmic electrode metals with a lower oxygen chemical potential would decrease the chance to generate excessive oxygen vacancies during cycling. Therefore, less RESET failures occur and a higher endurance is achieved. The theoretical findings of Guo et al. [7] were verified in this work by demonstrating increasing endurance going from Hf and Zr over Ti towards Ta as ohmic electrode metals. It was furthermore demonstrated that the same trend applies for the forming voltage. This results in a trade-off for the material selection between achievable endurance and required voltage to electroform the devices. This provides valuable information for future device engineering.

7.4 Outlook

Based on the presented results, it is suggested to keep the focus on device statistics instead of single cell characterization. The graphical tests of normality and log-normality used in this work [197, 198] should be extended by additional statistical tests as Chi-Squared test [199], Shapiro–Wilk test [200] or Jarque-Bera test, in order to deepen the understanding of the underlying distributions. Future studies could investigate in more detail how the intrinsic statistics could be tuned in favor of the application. It is evident that the internal statistics are linked closely to the dominating electrical conduction mechanism. A systematic variation of material systems and crystal structures could give rise to factors of influence. Furthermore, the cell size could be scaled down in combination with oxygen rich or poor environments to investigate the impact of a confined conducting filament.
Moreover, it was demonstrated that R2R fluctuations are initially very large. They decay exponentially and approach a rather constant level [86, 121]. Although the

fluctuations do not disappear entirely, a way to accelerate this decrease might prove to be valuable. Chang et al. report that soft RESET pulses during SET programming decrease read disturb failures [221]. Future studies could investigate if RESET algorithms containing both voltage polarities could produce more stable HRS distributions.

For both variability and retention, the statistical model presented in this work proved to be an excellent way to model the highly stochastic character of VCM ReRAM. The dynamical version of the model consistently explained both R2R variability and retention by introducing regions of different energy barriers for oxygen vacancy diffusion. Though, the defined regions are rather arbitrary in the current version. However, Schie et al. [212] suggested that small regions with low energy barriers confined by higher energy barriers are likely to be present in HfO_2. Therefore, a refinement of the presented model with multiple sub-diffusive regions seems reasonable and is very likely to provide an even better description of the observed short- and long-term instabilities.

With respect to the device endurance, the presented programming algorithm could be developed further. Yonai et al. reported that hard verify programming during the first 1000 cycles, followed by softer programming at higher cycle numbers resulted in the highest endurance [214]. Following these results, the algorithm presented in this work could be adapted to apply initially hard SET and RESET pulses. After the first few cycles it could return to the currently implemented goal of finding the lowest SET and RESET voltages which enable reliable switching.

Additionally, the study regarding the influence of the ohmic electrode could be extended towards metals with a higher oxygen chemical potential than Ta. In theory, this should provide even higher endurance. However, the trade-off with also increasing forming voltage has to be considered.

All in all, the key statement of this dissertation is that reliability aspects of VCM ReRAM have to be evaluated for their intrinsic statistics. Degradation effects like broadening of a read current (or resistance) distribution which are identified as key challenge could not have been detected on single or few devices. With respect to the large scale application of VCM ReRAM, the worst cells in a memory array are the ones which determine the reliability. To predict the behavior of these cells, it is crucial to understand the statistics of the whole system.

List of Tables

List of Abbreviations

1T1R	**1** **T**ransistor **1** **R**eRAM element
ADC	**A**nalog **D**igital **C**onverter
AE	**A**ctive **E**lectrode
AWG	**A**rbitrary **W**aveform **G**enerator
C2C	**C**ycle to **C**ycle
CBRAM	**C**onductive **B**ridge **R**andom **A**ccess **M**emory
CDF	**C**umulative **D**istribution **F**unction
CMOS	**C**omplementary **M**etal **O**xide **S**emiconductor
CRS	**C**omplementary **R**esistive **S**witching
D2D	**D**evice to **D**evice
DAC	**D**igital **A**nalog **C**onverter
DD	**D**rift-**D**iffusion
DFE	**D**efect **F**ormation **E**nergy
DRAM	**D**ynamic **R**andom **A**ccess **M**emory
DUT	**D**evice **U**nder **T**est
ECM	**E**lectro-**C**hemical **M**etallization memory
FHMM	**F**actorial **H**idden **M**arkov **M**odel
HF	**H**igh **F**requency
HMM	**H**idden **M**arkov **M**odel
LNS	**L**og **N**ormal **S**tatistics
LRS	**L**ow **R**esistive **S**tate
HRS	**H**igh **R**esistive **S**tate
KMC	**K**inetic **M**onte **C**arlo
MOSFET	**M**etal **O**xide **S**emiconductor **F**ield **E**ffect **T**ransistor
MRAM	**M**agnetoresistive **R**andom **A**ccess **M**emory
MTJ	**M**agnetic **T**unnel **J**unction

MTTF	**M**ean **T**ime **T**o **F**ailure
NVM	**N**on **V**olatile **M**emory
OE	**O**hmic **E**lectrode
OxRAM	**Ox**ide based **R**andom **A**ccess **M**emory
PCM	**P**hase **C**hange **M**emory
PDF	**P**robability **D**ensity **F**unction
PMC	**P**rogrammable **M**etallization **C**ell
QPC	**Q**uantum **P**oint **C**ontact
R2R	**R**ead to **R**ead
RAM	**R**andom **A**ccess **M**emory
ReRAM	**Re**dox based resistive switching **R**andom **A**ccess **M**emory
RRAM	**R**esistive switching **R**andom **A**ccess **M**emory
RTN	**R**andom **T**elegraph **N**oise
SEM	**S**canning **E**lectron **M**icroscopy
TAT	**T**rap **A**ssisted **T**unneling
TCM	**T**hermo-**C**hemical **M**emory
UV	**U**ltra **V**iolet
VCM	**V**alence **C**hange **M**emory

Bibliography

[1] G. E. Moore. “Cramming more components onto integrated circuits”. In: *Electronics* 38.8 (1965).

[2] I. L. Markov. “Limits on fundamental limits to computation”. In: *Nature* 512.7513 (2014), pp. 147–154.

[3] E. Abbaspour, S. Menzel, and C. Jungemann. “The Role of the Interface Reactions in the Electroforming of Redox-based Resistive Switching Devices Using KMC Simulations”. In: *(SISPAD)*. 2015, pp. 293–296.

[4] E. Abbaspour, S. Menzel, and C. Jungemann. “KMC Simulation of the Electroforming, Set and Reset Processes in Redox-based Resistive Switching Devices”. In: (SISPAD), 2016, pp. 141–144.

[5] E. Abbaspour, S. Menzel, and C. Jungemann. “Random telegraph noise analysis in Redox-based Resistive Switching Devices Using KMC Simulations (talk)”. In: (SISPAD), 2017.

[6] E. Abbaspour et al. “KMC Simulation of the Electroforming, Set and Reset Processes in Redox-based Resistive Switching Devices”. In: *IEEE Trans. Nanotechnol.* 17.6 (2018), pp. 1181–1188.

[7] Y. Guo and J. Robertson. “Material selection for oxide-based resistive random access memories”. In: *Appl. Phys. Lett.* 105.22 (2014), p. 223516.

[8] M. Irvine. “Early digital computers at Bell Telephone Laboratories”. In: *IEEE Ann. Hist. Comput.* 23.3 (2001), pp. 22–42.

[9] Y. Chen. “ReRAM: History, Status, and Future”. In: *IEEE Trans. Electron Devices* (2020), pp. 1–14.

[10] R. Waser, R. Bruchhaus, and S. Menzel. “Redox-based Resistive Switching Memories”. In: *Nanoelectronics and Information Technology (3rd edition)*. Ed. by R. Waser. Wiley-VCH, 2012, pp. 683–710.

[11] D. Ielmini and R. Waser. *Resistive Switching - From Fundamentals of Nanoionic Redox Processes to Memristive Device Applications*. Wiley-VCH, 2016.

[12] A. Sawa. “Resistive switching in transition metal oxides”. In: *Materials Today* 11.6 (2008), pp. 28–36.

[13] R. Waser et al. “Redox-Based Resistive Switching Memories - Nanoionic Mechanisms, Prospects, and Challenges”. In: *Adv. Mater.* 21.25-26 (2009), pp. 2632–2663.

[14] H.-S. P. Wong et al. “Metal–Oxide RRAM”. In: *Proc. IEEE* 100.6 (2012), pp. 1951–1970.

[15] M. A. Zidan, J. P. Strachan, and W. D. Lu. “The future of electronics based on memristive systems”. In: *Nat. Electron.* 1 (2018), pp. 22–29.

[16] M.-J. Lee et al. “A fast, high-endurance and scalable non-volatile memory device made from asymmetric Ta_2O_{5-x}/TaO_{2-x} bilayer structures”. In: *Nat. Mater.* 10.8 (2011), pp. 625–630.

[17] M. Witzleben et al. “Study of the SET switching event of VCM-based memories on a picosecond timescale”. In: *J. Appl. Phys.* 127.20 (2020), p. 204501.

[18] T. W. Hickmott. “Low-frequency negative resistance in thin anodic oxide films”. In: *J. Appl. Phys.* 33 (1962), pp. 2669–2682.

[19] J. F. Gibbons and W. E. Beadle. “Switching properties of thin NiO films”. In: *Solid-State Electron.* 7.11 (1964), pp. 785–790.

[20] W. R. Hiatt and T. W. Hickmott. “Bistable switching in niobium oxide diodes”. In: *Appl. Phys. Lett.* 6 (1965), pp. 106–108.

[21] A. Asamitsu et al. “Current switching of resistive states in magnetoresistive manganites”. In: *Nature* 388.6637 (1997), pp. 50–2.

[22] M. N. Kozicki et al. “Applications of programmable resistance changes in metal-doped chalcogenides”. In: *Proceedings of the International Solid-State Ionic Devices Conf., Seattle, WA, USA, 1999*. Center for Solid State Electron Res, Arizona State Univ, Tempe, AZ, USA. Electrochem. Soc, 1999, pp. 298–309.

[23] A. Beck et al. “Reproducible switching effect in thin oxide films for memory applications”. In: *Appl. Phys. Lett.* 77.1 (2000), pp. 139–41.

[24] R. Waser and M. Aono. “Nanoionics-based resistive switching memories”. In: *Nat. Mater.* 6.11 (2007), pp. 833–840.

[25] K. Gopalakrishnan et al. “Highly-Scalable Novel Access Device based on Mixed Ionic Electronic Conduction (MIEC) Materials for High Density Phase Change Memory (PCM) Arrays”. In: *Symposium on VLSI Technology (VLSIT), Honolulu, HI*. Symposium on VLSI Technology, 2010, pp. 205–206.

[26] Y.-B. Kim et al. “Bi-layered RRAM with Unlimited Endurance and Extremely Uniform Switching”. In: *IEEE Symp. VLSI Techn.* (2011), p. 52.

[27] S.-G. Park et al. "A Non-Linear ReRAM Cell with sub-1µA Ultralow Operating Current for High Density Vertical Resistive Memory (VRRAM)". In: *2012 International Electron Device Meeting*, 2012 International Electron Device Meeting, 2012.

[28] S. Sills et al. "A copper ReRAM cell for Storage Class Memory applications". In: *2014 Symposium on VLSI Technology (VLSI-Technology): Digest of Technical Papers*. 2014, pp. 1–2.

[29] J. Jameson et al. "Conductive-bridge memory (CBRAM) with excellent high-temperature retention". In: *Electron Devices Meeting (IEDM), 2013 IEEE International*. Electron Devices Meeting (IEDM), 2013 IEEE International, 2013, pp. 738–741.

[30] S. Fukuyama et al. "Suppression of endurance-stressed data-retention failures of 40nm TaOx-based ReRAM". In: IEEE (IRPS), 2018, P-MY.4-1-P-MY.4–5.

[31] E. Wu et al. "Filamentary Statistical Evolution from Nano-Conducting Path to Switching-Filament for Oxide-RRAM in Memory Applications". In: IEEE (IEDM), 2019, pp. 30.3.1–3.4.

[32] Y. Du et al. "Exploring the Impact of Random Telegraph Noise-Induced Accuracy Loss on Resistive RAM-Based Deep Neural Network". In: *IEEE Trans. Electron Devices* 67 (2020), pp. 3335–3340.

[33] Y. Lin et al. "Bayesian Neural Network Realization by Exploiting Inherent Stochastic Characteristics of Analog RRAM". In: *65th IEEE Annual International Electron Devices Meeting (IEDM), San Francisco, CA*. 2019 Ieee International Electron Devices Meeting (iedm), 2019.

[34] Y. Park, M. Kim, and J. Lee. "Emerging memory devices for artificial synapses". In: *J. Mater. Chem. C* 8.27 (2020), pp. 9163–9183.

[35] A. Valentian et al. "Fully Integrated Spiking Neural Network with Analog Neurons and RRAM Synapses". In: *65th IEEE Annual International Electron Devices Meeting (IEDM), San Francisco, CA*. 2019 Ieee International Electron Devices Meeting (iedm), 2019.

[36] C. Chen. "Understanding and Improving Reliability of Oxide-based Resistive RAM for Embedded Application and Storage Class Memory". PhD thesis. 2017.

[37] R. Strenz. "Review and Outlook on Embedded NVM Technologies – From Evolution to Revolution". In: *2020 IEEE International Memory Workshop (IMW)*. 2020, pp. 1–4.

[38] L.O. Chua. "Memristor-the missing circuit element". In: *IEEE Trans. Circuit Theory* CT-18.5 (1971), pp. 507–519.

[39] L.O. Chua and S.M. Kang. "Memristive devices and systems". In: *Proc. IEEE* 64.2 (1976), pp. 209–223.

[40] L.O. Chua. "Resistance switching memories are memristors". In: *Appl. Phys. A-Mater. Sci. Process.* 102.4 (2011), pp. 765–783.

[41] B. Dieny et al. "Magnetic Randon Access Memories". In: *Nanoelectronics and Information Technology (3rd ed.)* Ed. by R. Waser. Wiley-VCH, 2012, pp. 655–668.

[42] S. Raoux and M. Wuttig. "Information Storage Based on Phase Change Materials". In: *Nanoelectronics and Information Technology (3rd ed.)* Ed. by R. Waser. Wiley-VCH, 2012, pp. 669–682.

[43] S. Tappertzhofen. "Redox and Mass Transport Phenomena in Resistively Switching Thin Films". PhD thesis. 2014.

[44] R. Waser. "Resistive non-volatile memory devices (Invited Paper)". In: *16th Biennial Conference on Insulating Films on Semiconductors, Cambridge, ENGLAND*. Microelectronic Engineering, 2009, pp. 1925–1928.

[45] L. Goux and S. Spiga. "Unipolar Resistive-Switching Mechanisms". In: 1st ed. Wiley-VCH, 2016. Chap. 13, pp. 363–393.

[46] D. Ielmini, R. Bruchhaus, and R. Waser. "Thermochemical resistive switching: materials, mechanisms, and scaling projections". In: *Phase Transit.* 84.7 (2011), pp. 570–602.

[47] R. Waser. "Electrochemical and Thermochemical Memories". In: *IEDM*. IEEE International Electron Devices Meeting 2008, Technical Digest, 2008, pp. 289–292.

[48] D. Ielmini, F. Nardi, and C. Cagli. "Universal Reset Characteristics of Unipolar and Bipolar Metal-Oxide RRAM". In: *Transactions on Electron Devices* 58.10 (2011), pp. 1–8.

[49] M. Lübben et al. "Graphene-Modified Interface Controls Transition from VCM to ECM Switching Modes in Ta/TaOx Based Memristive Devices". In: *Adv. Mater.* 27.40 (2015), pp. 6202–6207.

[50] M. Luebben et al. "SET kinetics of electrochemical metallization cells - Influence of counter electrodes in SiO_2/Ag based systems". In: *Nanotechnology* 28.13 (2017), pp. 135205/1–6.

[51] M. Lübben and I. Valov. "Active Electrode Redox Reactions and Device Behavior in ECM Type Resistive Switching Memories". In: *Adv. El. Mater.* 5.9 (2019), p. 1800933.

[52] I. Valov et al. "Electrochemical metallization memories-fundamentals, applications, prospects". In: *Nanotechnology* 22.25 (2011), pp. 254003/1–22.

[53] M. N. Kozicki, M. Mitkova, and I.Valov. "Electrochemical Metallization Memories". In: ed. by R. Waser; D. Ielmini. Wiley-VCH Verlag GmbH & Co. KGaA, 2016. Chap. 17, pp. 483–514.

[54] M. N. Kozicki and H. J. Barnaby. "Conductive bridging random access memory-materials, devices and applications". In: *Semiconductor Science and Technology* 31.11 (2016), pp. 113001 / 1–32.

[55] L. Goux and I. Valov. "Electrochemical processes and device improvement in conductive bridge RAM cells". In: *Phys. Status Solidi A* 213 (2016), pp. 274–288.

[56] E. Linn et al. "Complementary Resistive Switches for Passive Nanocrossbar Memories". In: *Nat. Mater.* 9.5 (2010), pp. 403–406.

[57] F. Nardi et al. "Complementary switching in metal oxides: toward diodeless crossbar RRAMs". In: *2011 IEEE International Electron Devices Meeting - IEDM '11*. Ed. by IEEE International Electron Devices Meeting (IEDM). IEEE International Electron Devices Meeting (IEDM), 2011, pp. 31.1.1–31.1.4.

[58] A. Schoenhals et al. "Critical ReRAM Stack Parameters Controlling Complementary versus Bipolar Resistive Switching". In: *Memory Workshop (IMW), 2015 IEEE International*. Memory Workshop (IMW), 2015 IEEE International, 2015, pp. 73–76.

[59] C. La Torre et al. "Compact Modeling of Complementary Switching in Oxide-Based ReRAM Devices". In: *IEEE Trans. Electron Devices* 66.3 (2019), pp. 1268–1275.

[60] R. Waser et al. "Introduction to Nanoionic Elements for Information Technology". In: ed. by D. Ielmini; R. Waser. Wiley-VCH, 2016. Chap. 1, pp. 1–29.

[61] F. A. Kroeger. *The chemistry of imperfect crystals*. North-Holland, 1973.

[62] D. S. Jeong, B. J. Choi, and C. S. Hwang. "Resistive Switching - Electroforming Processes in Metal Oxide Resistve-Switching Cells". In: ed. by D. Ielmini; R. Waser. Wiley-VCH, 2016. Chap. 10, pp. 289–316.

[63] M. Lübben et al. "Processes and Effects of Oxygen and Moisture in Resistively Switching TaOx and HfOx". In: *Adv. Electr. Mat.* 4.1 (2018), p. 1700458.

[64] J. J. Yang et al. "The mechanism of electroforming of metal oxide memristive switches". In: *Nanotechnology* 20.21 (2009), p. 215201.

[65] G. Cellere et al. "Influence of dielectric breakdown on MOSFET drain current". In: *IEEE transactions on electron devices* 52.2 (2005), pp. 211–217.

[66] M. Noman et al. "Transient characterization of the electroforming process in TiO2 based resistive switching devices". In: *Appl. Phys. Lett.* 102.2 (2013).

[67] T. Menke et al. "Impact of the electroforming process on the device stability of epitaxial Fe-doped $SrTiO_3$ resistive switching cells". In: *J. Appl. Phys.* 106.11 (2009), p. 114507.

[68] D. Cooper et al. "Anomalous Resistance Hysteresis in Oxide ReRAM: Oxygen Evolution and Reincorporation Revealed by in situ TEM". In: *Adv. Mater.* 29.23 (2017), p. 1700212.

[69] F. Miao et al. "Observation of two resistance switching modes in TiO_2 memristive devices electroformed at low current". In: *Nanotechnology* 22.11 (2011), pp. 254007/1–7.

[70] X. Sun et al. "Bipolar resistance switching characteristics with opposite polarity of Au/$SrTiO_3$/Ti memory cells". In: *Nanoscale Res. Lett.* 6 (2011), pp. 599/1–8.

[71] A. Schönhals et al. "Role of the Electrode Material on the RESET Limitation in Oxide ReRAM Devices". In: *Adv. Electron. Mater.* 4.2 (2017), pp. 1700243/1–11.

[72] H. Zhang et al. "Resistive Switching Modes with Opposite Polarity in Pt/TiO_2-/Ti/Pt Nano-sized ReRAM Devices". In: *ACS Applied Materials Interfaces* 10.35 (2018), pp. 29766–29778.

[73] C. La Torre. "Physics-Based Compact Modeling of Valence-Change-Based Resistive Switching Devices". PhD thesis. 2019.

[74] R. Degraeve et al. "Dynamic Hour Glass Model for SET and RESET in HfO_2 RRAM". In: *Proceedings of the 2012 Symposium on VLSI Technology* (2012), p. 75.

[75] L. M. Procel et al. "Experimental evidence of the quantum point contact theory in the conduction mechanism of bipolar HfO2-based resistive random access memories". In: *J. Appl. Phys.* 114.7 (2013), pp. 74509/1–.

[76] E. Miranda et al. "Multi-channel conduction in redox-based resistive switch modelled using quantum point contact theory". In: *Applied Physics Letters* 103.22 (2013), pp. 222904/1–5.

[77] X. Lian et al. "Multi-scale quantum point contact model for filamentary conduction in resistive random access memories devices". In: *J. Appl. Phys.* 115.24 (2014), 244507 (8 pp.)–24450.

[78] S. Yu, X. Guan, and H. P. Wong. "Conduction mechanism of TiN/HfOx/Pt resistive switching memory: A trap-assisted-tunneling model". In: *Appl. Phys. Lett.* 99.6 (2011), pp. 063507–063507.

[79] D. Garbin et al. "Resistive memory variability: A simplified trap-assisted tunneling model". In: *SOLID STATE ELECTRON* 115 (2016), pp. 126–132.

[80] T. Zanotti, F. M. Puglisi, and P. Pavan. "Circuit Reliability of Low-Power RRAM-Based Logic-in-Memory Architectures". In: IEEE (IIRW), 2019, pp. 1–5.

[81] P. K. Radtke et al. "Stochastic dynamics of resistive switching: fluctuations lead to optimal particle number". In: *New Journal of Physics* 19.9 (Sept. 2017), p. 093007.

[82] D. J. Wouters et al. "Reliability Aspects". In: *Resistive switching. From Fundamentals of Nanoionic Redox Processes to Memristive Device Applications*. Ed. by D. Ielmini; R. Waser. Wiley, 2016.

[83] A. Kindsmüller et al. "On the role of the metal oxide/reactive electrode interface during the forming procedure of valence change ReRAM devices". In: *Nanoscale* 11 (2019), pp. 18201–18208.

[84] C. Zambelli et al. "Statistical analysis of resistive switching characteristics in ReRAM test arrays". In: *2014 International Conference on Microelectronic Test Structures (ICMTS)*. 2014, pp. 27–31.

[85] A. Fantini et al. "Intrinsic Switching Variability in HfO2 RRAM". In: IEEE (IMW), 2013, pp. 30–33.

[86] A. Fantini et al. "Intrinsic program instability in HfO2 RRAM and consequences on program algorithms". In: IEEE (IEDM), 2015, pp. 7–5.

[87] C. Y. Chen et al. "Novel Flexible and Cost-Effective Retention Assessment Method for TMO-Based RRAM". In: *IEEE Electron Device Lett.* 37.9 (2016), pp. 1112–1115.

[88] S. Ambrogio et al. "Statistical Fluctuations in HfOx Resistive-Switching Memory: Part I - Set/Reset Variability". In: *IEEE Trans. Electron Devices* 61 (2014), pp. 2912–2919.

[89] E. Perez et al. “Reduction of the Cell-to-Cell Variability in Hf1-xAlxOy Based RRAM Arrays by Using Program Algorithms”. In: *IEEE Electron Device Lett.* 38.2 (2017), pp. 175–178.

[90] A. Schoenhals, R. Waser, and D. Wouters. “Improvement of SET Variability in TaOx Based Resistive RAM Devices”. In: *Nanotechnology* 28 (2017), p. 465203.

[91] K. Fleck. “The Influence of Local Heating on the Switching Kinetics of Resistive Switching Oxides”. PhD thesis. 2018.

[92] S. Leisten. “Influence of the ohmic electrode on the switching kinetics and endurance of ZrO2-based resisitive memory cells”. Bachelor thesis, RWTH Aachen University. 2018.

[93] M. Hüttemann. “Optimizing the Endurance of Resistive Switching VCM Memory Cells”. Bachelor thesis, RWTH Aachen University. 2019.

[94] A. Prakash et al. “Resistance controllability and variability improvement in a TaOx-based resistive memory for multilevel storage application”. In: *Appl. Phys. Lett.* 106.23 (2015), pp. 233104/1–4.

[95] C. La Torre et al. “Dependence of the SET switching variability on the initial state in HfOx-based ReRAM”. In: *Phys. Status Solidi A* 213.2 (2016), pp. 316–319.

[96] N. Raghavan et al. “Stochastic variability of vacancy filament configuration in ultra-thin dielectric RRAM and its impact on OFF-state reliability”. In: *Electron Devices Meeting (IEDM), 2013 IEEE International*. Electron Devices Meeting (IEDM), 2013 IEEE International, 2013, pp. 21.1.1–21.1.4.

[97] F.M. Puglisi et al. “Instability of HfO 2 RRAM devices: comparing RTN and cycling variability”. In: *2014 IEEE International Reliability Physics Symposium (IRPS)*. 2014 IEEE International Reliability Physics Symposium (IRPS), 2014, MY.5.1–MY.5.5.

[98] F. Puglisi, A. Qafa, and P. Pavan. “Temperature Impact on the Reset Operation in HfO 2 RRAM”. In: *IEEE Electron Device Lett.* 36.3 (2015), pp. 244–6.

[99] K. M. Kim et al. “Voltage divider effect for the improvement of variability and endurance of TaOx memristor”. In: *Sci Rep* 6 (2016), pp. 20085/1–6.

[100] W. Banerjee et al. “Variability Improvement of TiOx/Al2O3 Bilayer Nonvolatile Resistive Switching Devices by Interfacial Band Engineering with an Ultrathin Al2O3 Dielectric Material”. In: *ACS Omega* 2.10 (2017), pp. 6888–6895.

[101] C. Nguyen et al. “Study of Forming impact on 4k Bit RRAM array performances and reliability”. In: *International Memory Workshop 2017, Monterey, CA, USA*. International Memory Workshop 2017, 2017.

[102] X. Li et al. “Short Time High-Resistance State Instability of TaOx-Based RRAM Devices”. In: *IEEE Electron Device Lett.* 38.1 (2017), pp. 32–35.

[103] Y. Lin et al. “A comprehensive study of 3-stage high resistance state retention behavior for TMO ReRAMs from single cells to a large array”. In: IEEE (IEDM), 2017, pp. 2.5.1–2.5.4.

[104] T. Gong et al. “Classification of Three-Level Random Telegraph Noise and Its Application in Accurate Extraction of Trap Profiles in Oxide-Based Resistive Switching Memory”. In: *IEEE Electron Device Lett.* 39.9 (2018), pp. 1302–1305.

[105] P. Huang et al. “RTN based oxygen vacancy probing method for Ox-RRAM reliability characterization and its application in tail bits”. In: IEEE (IEDM), 2017, pp. 21.4.1–21.4.4.

[106] F. M. Puglisi et al. “Random Telegraph Noise in Resistive Random Access Memories: Compact Modeling and Advanced Circuit Design”. In: *IEEE Trans. Electron Devices* 65.7 (2018), pp. 2964–2972.

[107] S. Balatti et al. “Voltage-dependent random telegraph noise (RTN) in HfOx resistive RAM”. In: *2014 IEEE International Reliability Physics Symposium*. 2014 IEEE International Reliability Physics Symposium, 2014.

[108] F.M. Puglisi et al. “A Complete Statistical Investigation of RTN in HfO2-Based RRAM in High Resistive State”. In: *IEEE Transactions on Electronic Devices* 62.8 (2015), pp. 2606–2613.

[109] D. Veksler et al. “Methodology for the statistical evaluation of the effect of random telegraph noise (RTN) on RRAM characteristics”. In: *IEEE Electron Device Lett.* 33.12 (2012).

[110] N. Raghavan et al. “RTN Insight to Filamentary Instability and Disturb Immunity in Ultra-Low Power Switching HfO_x and AlO_x”. In: *Proceedings of the 2013 Symposium on VLSI Technology* (2013).

[111] S. Brivio et al. “Stimulated Ionic Telegraph Noise in Filamentary Memristive Devices”. In: *SCIENTIFIC REPORTS* 9 (2019).

[112] S. Lee et al. “Characterization of oxide traps leading to RTN in high-k and metal gate MOSFETs”. In: *2009 IEEE International Electron Devices Meeting (IEDM)*. 2009, pp. 1–4.

[113] M.B. da Silva andH. Tuinhout et al. "A physics-based RTN variability model for MOSFETs". In: *2014 IEEE International Electron Devices Meeting*. 2014, pp. 35.2.1–35.2.4.

[114] G. Kapila and V. Reddy. "Impact of Sampling Rate on RTN Time Constant Extraction and Its Implications on Bias Dependence and Trap Spectroscopy". In: *IEEE Trans. Device Mater. Reliab.* 14 (2014), pp. 616–622.

[115] S. Ambrogio et al. *Noise-induced resistance broadening in resistive switching memory - Part II: array statistics*. 2015.

[116] D. Ielmini, F. Nardi, and C. Cagli. "Resistance-dependent amplitude of random telegraph-signal noise in resistive switching memories". In: *Applied Physics Letters* 96.5 (2010), pp. 53503 / 1–3.

[117] S. Ambrogio et al. "Data retention statistics and modelling in HfO2 resistive switching memories". In: *Reliability Physics Symposium (IRPS), 2015 IEEE International*. Reliability Physics Symposium (IRPS), 2015 IEEE International, 2015, MY.7.1–MY.7.6.

[118] A. Fantini et al. "Intrinsic switching behavior in HfO_2 RRAM by fast electrical measurements on novel 2R test structures". In: *Proc. 4th IEEE Int. Memory Workshop* (2012), pp. 1–4.

[119] Z. Chai et al. "Probing the Critical Region of Conductive Filament in Nanoscale HfO2 Resistive-Switching Device by Random Telegraph Signals". In: *IEEE Trans. Electron Devices* 64.10 (2017), pp. 4099–4105.

[120] J. Ma et al. "Investigation of Preexisting and Generated Defects in Nonfilamentary a-Si/TiO2 RRAM and Their Impacts on RTN Amplitude Distribution". In: *IEEE Trans. Electron Devices* 65.3 (2018), pp. 970–977.

[121] S. Clima et al. "Intrinsic Tailing of Resistive States Distributions in Amorphous HfOx and TaOx Based Resistive Random Access Memories". In: *Electron Device Letters, IEEE* 36 (2015), pp. 769–771.

[122] N. Agudov et al. "Nonstationary distributions and relaxation times in a stochastic model of memristor". In: *Journal of Statistical Mechanics: Theory and Exper* 2020.2 (2020), p. 024003.

[123] V. G. Karpov and D. Niraula. "Log-Normal Statistics in Filamentary RRAM Devices and Related Systems". In: *IEEE Electron Device Lett.* 38 (2017), pp. 1240–1243.

[124] A. Benoist et al. "28nm advanced CMOS resistive RAM solution as embedded non-volatile memory". In: *2014 IEEE International Reliability Physics Symposium*. 2014, 2E.6.1–2E.6.5.

[125] S. Yu et al. "Characterization of low-frequency noise in the resistive switching of transition metal oxide HfO_2". In: *Phys. Rev. B: Condens. Matter* 85.4 (2012), pp. 45324 / 1–5.

[126] S. B. Lee et al. "Large 1/f noise of unipolar resistance switching and its percolating nature". In: *Appl. Phys. Lett.* 95.12 (2009), pp. 122112/1–.

[127] M. Stoyanov, M. Gunzburger, and J. Burkardt. "Pink Noise, $1/f^{\alpha}$ Noise, and their Effect on Solutions of Differential Equations". In: *International Journal for Uncertainty Quantification* 1.3 (2011), pp. 257–278. ISSN: 2152-5080.

[128] F. M. Puglisi and P. Pavan. "RTN analysis with FHMM as a Tool for Multi-Trap Characterization in HfOx RRAM". In: *2013 IEEE International Conference of Electron Devices and Solid-State Circuits*. IEEE. IEEE, 2013, pp. 1–2.

[129] Y. Chen et al. "Endurance/Retention Trade-off on HfO2\Metal Cap 1T1R Bipolar RRAM". In: *IEEE Trans. Electron Devices* 60.3 (2013), pp. 1114–1121.

[130] Y. Y. Chen et al. "Improvement of data retention in HfO2/Hf 1T1R RRAM cell under low operating current". In: *Electron Devices Meeting (IEDM), 2013 IEEE International*. Electron Devices Meeting (IEDM), 2013 IEEE International, 2013, pp. 10.1.1–10.1.4.

[131] Y. Y. Chen et al. "Understanding of the Endurance Failure in Scaled HfO2-based 1T1R RRAM through Vacancy Mobility Degradation". In: *2012 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA*. 2012 IEEE International Electron Devices Meeting (IEDM), 2012, pp. 20.3.1–20.3.4.

[132] Z. Wei et al. "Demonstration of high-density ReRAM ensuring 10-year retention at 85°C based on a newly developed reliability model". In: *Technical Digest - International Electron Devices Meeting, IEDM*. Technical Digest - International Electron Devices Meeting, IEDM, 2011, pp. 31.4.1–31.4.4.

[133] Z. Wei et al. "Retention model for high-density ReRAM". In: *4th IEEE International Memory Workshop (IMW)*. 4th IEEE International Memory Workshop (IMW), 2012.

[134] J. Frascaroli et al. "Effect of Al doping on the retention behavior of memories HfO2 resistive switching". In: *Microelectron. Eng.* 147 (2015), pp. 104–107.

[135] D. Ielmini et al. "Size-Dependent Retention Time in NiO-Based Resistive-Switching Memories". In: *IEEE Electron Device Lett.* 31.4 (2010), pp. 353–355.

[136] S. Yu et al. "A Monte Carlo study of the low resistance state retention of HfOx based resistive switching memory". In: *Appl. Phys. Lett.* 100.4 (2012), pp. 43507/1–.

[137] E. Perez et al. “Data retention investigation in Al:HfO2-based resistive random access memory arrays by using high-temperature accelerated tests”. In: *Journal of Vacuum Science & Technology B* 37.1 (2019), p. 012202.

[138] B. Traoré et al. “On the Origin of Low-Resistance State Retention Failure in HfO2-Based RRAM and Impact of Doping/Alloying”. In: *IEEE Trans. Electron Devices* 62 (2015), pp. 4029–4036.

[139] M. Azzaz et al. “Endurance/Retention Trade Off in HfOx and TaOx Based RRAM”. In: IEEE (IMW), 2016, pp. 1–4.

[140] Y. Wang et al. “Algorithm-Enhanced Retention Based on Megabit Array of Cu_xSi_yO RRAM”. In: *IEEE Electron Device Lett.* 33.10 (2012), pp. 1408–1410.

[141] S. Fukuyama et al. “Comprehensive Analysis of Data-Retention and Endurance Trade-Off of 40nm TaOx-based ReRAM”. In: IEEE (IRPS), 2019, pp. 1–6.

[142] T. Ninomiya et al. “Improvement of Data Retention During Long-Term Use by Suppressing Conductive Filament Expansion in TaOx Bipolar-ReRAM”. In: *IEEE Electron Device Lett.* 34.6 (2013), pp. 762–764.

[143] B. Govoreanu et al. “10x10 nm2 Hf/HfOx Crossbar Resistive RAM with Excellent Performance, Reliability and Low-Energy Operation”. In: *2011 IEEE International Electron Devices Meeting - IEDM ’11*. IEDM Tech. Dig., 2011, pp. 31.6.1–31.6.4.

[144] P. Huang et al. “Analytic Model for Statistical State Instability and Retention Behaviors of Filamentary Analog RRAM Array and Its Applications in Design of Neural Network”. In: IEEE (IEDM), 2018, pp. 40.4.1–40.4.4.

[145] M. Zhao et al. “Investigation of statistical retention of filamentary analog RRAM for neuromophic computing”. In: IEEE (IEDM), 2017, pp. 39.4.1–39.4.4.

[146] J. Park et al. “Investigation of State Stability of Low-Resistance State in Resistive Memory”. In: *IEEE Electron Device Lett.* 31.5 (2010), pp. 485–487.

[147] S. Choi et al. “Retention failure analysis of metal-oxide based resistive memory”. In: *Appl. Phys. Lett.* 105 (2014), p. 113510.

[148] S. Arrhenius. “Über die Reaktionsgeschwindigkeit bei der Inversion von Rohrzucker durch Säuren”. In: *Zeitschrift für physikalische Chemie* 4.1 (1889), pp. 226–248.

[149] J. H. Hur. “First principles study of oxygen vacancy activation energy barrier in zirconia-based resistive memory”. In: *Sci Rep* 10.1 (2020), pp. 5405/1–8.

[150] H. Y. Lee et al. "Low Power and High Speed Bipolar Switching with A Thin Reactive Ti Buffer Layer in Robust HfO_2 Based RRAM". In: *IEEE International Electron Devices Meeting, San Francisco, CA*. IEEE International Electron Devices Meeting 2008, Technical Digest, 2008, pp. 297–300.

[151] H. Lee et al. "Evidence and solution of over-RESET problem for HfOX based resistive memory with sub-ns switching speed and high endurance". In: *Technical Digest - International Electron Devices Meeting, IEDM*. Technical Digest - International Electron Devices Meeting, IEDM, 2010, pp. 19.7.1–19.7.4.

[152] G. Sassine et al. "Optimizing Programming Energy for Improved RRAM Reliability for High Endurance Applications". In: *2018 IEEE International Memory Workshop (IMW)*. 2018 IEEE International Memory Workshop (IMW), 2018, pp. 1–4.

[153] Y. Chen et al. "Balancing SET/RESET Pulse for >1010 Endurance in HfO2 1T1R Bipolar RRAM". In: *IEEE Trans. Electron Devices* 59.12 (2012), pp. 3243–3249.

[154] J. J. Yang et al. "High switching endurance in TaO_x memristive devices". In: *Appl. Phys. Lett.* 97.23 (2010), pp. 232102/1–3.

[155] Z. Wei et al. "Highly Reliable TaOx ReRAM and Direct Evidence of Redox Reaction Mechanism". In: *IEEE Tech. Dig.* (2008).

[156] A. Kawahara et al. "An 8Mb multi-layered cross-point ReRAM macro with 443MB/s write throughput". In: *ISSCC Dig. Tech Papers* (2012), pp. 432–434.

[157] S. Koveshnikov et al. "Real-time study of switching kinetics in integrated 1T/HfOx 1R RRAM: Intrinsic tunability of set/reset voltage and trade-off with switching time". In: *Technical Digest - International Electron Devices Meeting, IEDM*. Technical Digest - International Electron Devices Meeting, IEDM, 2012, pp. 20.4.1–20.4.3.

[158] X. D. Huang et al. "Forming-Free, Fast, Uniform, and High Endurance Resistive Switching From Cryogenic to High Temperatures in W/AlOx/Al2O3/Pt Bilayer Memristor". In: *IEEE Electron Device Lett.* 41.4 (2020), pp. 549–552.

[159] C. W. Hsu et al. "Homogeneous barrier modulation of TaOx/TiO2 bilayers for ultra-high endurance three-dimensional storage-class memory". In: *Nanotechnology* 25.16 (2014), pp. 165202/1–7.

[160] Y. Meng et al. "Fast step-down set algorithm of resistive switching memory with low programming energy and significant reliability improvement". In: *VLSI-Technology: Digest of Technical Papers*. 2014, pp. 1–2.

[161] J. Muñoz-Gorriz et al. "Impact of the forming and cycling processes on the electrical and physical degradation characteristics of HfO2-based resistive switching devices". In: *Thin Solid Films* 706 (2020), p. 138027.

[162] M. Mao et al. "Programming strategies to improve energy efficiency and reliability of ReRAM memory systems". In: *2015 IEEE Workshop on Signal Processing Systems (SiPS)*. 2015, pp. 1–6.

[163] E. Perez et al. "Programming Pulse Width Assessment for Reliable and Low-Energy Endurance Performance in Al:HfO2-Based RRAM Arrays". In: *ELECTRONICS* 9 (2020).

[164] C. Nail et al. "Understanding RRAM endurance, retention and window margin trade-off using experimental results and simulations". In: *2016 IEEE International Electron Devices Meeting (IEDM)*. 2016 IEEE International Electron Devices Meeting (IEDM), 2016, pp. 4.5.1–4.5.4.

[165] K. Z. Rushchanskii, S. Bluegel, and M. Lezaic. "Routes for increasing endurance and retention in HfO2-based resistive switching memories". In: *Phys. Rev. Mater.* 2.11 (2018), pp. 115002/1–9.

[166] Y. Y. Chen et al. "Tailoring switching and endurance/retention reliability characteristics of HfO2/Hf RRAM with Ti, Al, Si dopants". In: *34th Symposium on VLSI Technology (VLSI-Technology): Digest of Technical Papers, Honolulu, HI*. 2014 Symposium On Vlsi Technology (vlsi-Technology): Digest of Technical Papers, 2014.

[167] C. Y. Chen et al. "Understanding the Impact of Programming Pulses and Electrode Materials on the Endurance Properties of Scaled Ta$_2$O$_5$ RRAM Cells". In: *2014 IEEE International Electron Devices Meeting (IEDM)*. 2014 IEEE International Electron Devices Meeting (IEDM), 2014, pp. 14.2.1–14.2.4.

[168] W. Kim et al. "Impact of oxygen exchange reaction at the ohmic interface in Ta_2O_5-based ReRAM devices". In: *Nanoscale* 8.41 (2016), pp. 17774–17781.

[169] L. R. Rabiner. "A tutorial on hidden Markov models and selected applications in speech recognition". In: *Proceedings of the IEEE* 77 (1989), pp. 257–286.

[170] Z. Ghahramani and M. Jordan. "Factorial hidden Markov models Machine Learning". In: *Kluwer Academic Publishers* (1997).

[171] F. M. Puglisi and P. Pavan. "Guidelines for a Reliable Analysis of Random Telegraph Noise in Electronic Devices". In: *IEEE Transactions on Instrumentation and Measureme* 65 (2016), pp. 1435–1442.

[172] AJ Viterbi. "Error bounds for convolutional codes and an asymptotically optimal decoding algorithm. IEEE Trans. Informat. Theory". In: (1967).

[173] J. R. Otterpohl. "Baum-Welch Learning in Discrete Hidden Markov Models with Linear Factorial Constraints". In: *Artificial Neural Networks — ICANN 2002*. Ed. by Springer Berlin Heidelberg. 2002, pp. 1180–1185.

[174] L. E. Baum et al. "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains". In: *Ann. Math. Statist.* 41.1 (1970), pp. 164–171.

[175] X. Guan, S. Yu, and H. Wong. "On the Switching Parameter Variation of Metal-Oxide RRAM-Part I: Physical Modeling and Simulation Methodology". In: *IEEE Trans. Electron Devices* 59.4 (2012), pp. 1172–1182.

[176] A. Padovani et al. "Microscopic Modeling of HfOx RRAM Operations: From Forming to Switching". In: *IEEE Trans. Electron Devices* 62.6 (2015), pp. 1998–2006.

[177] L. Larcher et al. "Microscopic understanding and modeling of HfO2 RRAM device physics". In: *2012 International Electron Devices Meeting*. 2012, pp. 20.1.1–20.1.4.

[178] A. Makarov, V. Sverdlov, and S. Selberherr. "Stochastic modeling of bipolar resistive switching in metal-oxide based memory by Monte Carlo technique". In: *Journal of Computational Electronics* 9.3 (2010), pp. 146–152.

[179] S. Aldana et al. "Resistive switching in HfO2 based valence change memories, a comprehensive 3D kinetic Monte Carlo approach". In: *J. Phys. D Appl. Phys.* 53.22 (2020), p. 225106.

[180] E. Abbaspour, S. Menzel, and C. Jungemann. "Studying the switching variability in redox-based resistive switching devices". In: *Journal of Computational Electronics* (2020).

[181] L. Lundkvist, I. Lundstrom, and C. Svensson. "Discharge of MNOS Structures". In: *Solid-State Electron.* 16.7 (1973), pp. 811+.

[182] M. Schie et al. "Field-enhanced route to generating anti-Frenkel pairs in HfO_2". In: *Physical Review Materials* 2.3 (2018), p. 035002.

[183] W. Stehling et al. "Kinetic Monte Carlo modeling of the charge transport in a HfO2-based ReRAM with a rough anode". In: *2017 17th Non-Volatile Memory Technology Symposium (NVMTS), 30 Aug.- 1. Sep. 2017, Aachen, Germany*. 2017 17th Non-Volatile Memory Technology Symposium (NVMTS), 30 Aug.- 1. Sep. 2017, Aachen, Germany, 2017.

[184] George G Judge et al. *Introduction to the Theory and Practice of Econometrics*. 330.015195 I61 1988. J. Wiley, 1988.

[185] J.R. Rumble and J. Rumble. *CRC Handbook of Chemistry and Physics, 98th Edition*. CRC Handbook of Chemistry and Physics. CRC Press LLC, 2017. ISBN: 9781498784542.

[186] K. Zhang et al. "Progress in rectifying-based RRAM passive crossbar array". In: *Sci. China-Technol. Sci.* 54.4 (2011), pp. 811–818.

[187] H. Li et al. "Looking Ahead for Resistive Memory Technology A broad perspective on ReRAM technology for future storage and computing". In: *IEEE Consum. Electron. Mag.* 6.1 (2017), pp. 94–103.

[188] A. Siemon et al. "A Complementary Resistive Switch-based Crossbar Array Adder". In: *IEEE J. Emerging Sel. Top. Circuits Syst.* 5.1 (2015), pp. 64–74.

[189] M. Mayahinia et al. "A Novel Voltage Controlled Oscillation based ADC Design for Computation-in-Memory Using Emerging ReRAMs". In: *Journal on Emerging Technologies in Computing Systems, accepted* (2021).

[190] J. Hur and D. Lee. "Universal Memory Characteristics and Degradation Features of ZrO2-Based Bipolar Resistive Memory". In: *ADVANCED ELECTRONIC MATERIALS* (2020).

[191] G. H. Kim et al. "Influence of the Interconnection Line Resistance and Performance of a Resistive Cross Bar Array Memory". In: *J. Electrochem. Soc.* 157.10 (2010), G211–G215.

[192] A. Schoenhals et al. "3-bit Read Scheme for Single Layer Ta_2O_5 ReRAM". In: *Non-Volatile Memory Technology Symposium (NVMTS), 2014 14th Annual*. Non-Volatile Memory Technology Symposium (NVMTS), 2014.

[193] K. Fleck et al. "Uniting Gradual and Abrupt SET Processes in Resistive Switching Oxides". In: *Phys. Rev. Applied* 6.6 (2016), p. 064015.

[194] S. Wiefels et al. "HRS Instability in Oxide based Bipolar Resistive Switching Cells". In: *IEEE Trans. Electron Devices* 67.10 (2020), pp. 4208–4215.

[195] E. Perez et al. "Reduction of the Cell-to-Cell Variability in $Hf_{1-x}Al_xO_y$ Based RRAM Arrays by Using Program Algorithms". In: *IEEE Electron Device Lett.* PP (2016), pp. 1–1.

[196] B. Yazici and S. Yolacan. "A comparison of various tests of normality". In: *Journal of Statistical Computation and Simulation* 77.2 (2007), pp. 175–183.

[197] K. R. Das and A. H. M. R. Imon. "A brief review of tests for normality". In: *American Journal of Theoretical and Applied Statistics* 5.1 (2016), pp. 5–12.

[198] M. C. Wang and B. J. Bushman. "Using the normal quantile plot to explore meta-analytic data sets." In: *Psychological Methods* 3.1 (1998), p. 46.

[199] K. Pearson F.R.S. "X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling". In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 50.302 (1900), pp. 157–175.

[200] S. S. Shapiro and M. B. Wilk. "An Analysis of Variance Test for Normality (Complete Samples)". In: *Biometrika* 52.3/4 (1965), pp. 591–611.

[201] C. M. Jarque and A. K. Bera. "Efficient tests for normality, homoscedasticity and serial independence of regression residuals". In: *Economics Letters* 6.3 (1980), pp. 255–259.

[202] T. Thadewald and H. Büning. *Jarque-Bera test and its competitors for testing normality: A power comparison*. Discussion Papers 9. Free University Berlin, School of Business & Economics, 2004.

[203] B. W. Yap and C. H. Sim. "Comparisons of various types of normality tests". In: *Journal of Statistical Computation and Simulation* 81.12 (2011), pp. 2141–2155.

[204] Y. R. Gel and J. L. Gastwirth. "A robust modification of the Jarque-Bera test of normality". In: *Economics Letters* 99.1 (2008), pp. 30–32.

[205] JART. *Juelich Aachen Resistive Switching Tools (JART)*. Tech. rep. 2019.

[206] C. Bengel et al. "Variability-Aware Modeling of Filamentary Oxide based Bipolar Resistive Switching Cells Using SPICE Level Compact Models". In: *TCAS 1* (2020).

[207] D.J. Wouters et al. "On the universality of I-V switching characteristics in non-volatile and volatile resistive switching oxides". In: *Faraday Discuss.* 213 (2019), pp. 183–196.

[208] N. Kopperberg. "Physical Modeling of Instability ans Retention Effects in VCM ReRAM Devices Using Kinetic Monte Carlo Methods". Master thesis, RWTH Aachen University. 2019.

[209] S. Wiefels et al. "Empirical Tunneling Model Describing the Retention of 2.5 Mb HfO2 based ReRAM". In: *International Symposium On VLSI Technology, Systems and Application (VLSI-TSA)*. IEEE, 2020.

[210] S. Wiefels et al. "Statistical Modeling and Understanding of HRS Retention in 2.5 Mb HfO2 based ReRAM". In: *IMW 2020* (2020), pp. 28–31.

[211] H. Schroeder and D.S. Jeong. "Resistive switching in a Pt/TiO_2/Pt thin film stack – a candidate for a non-volatile ReRAM". In: *Microelectronic Engineering* 84 (2007), pp. 1982–1985.

[212] M. Schie et al. “Ion migration in crystalline and amorphous HfOX”. In: *J. Chem. Phys.* 146.9 (2017), pp. 94508/1–9.

[213] S. Wiefels et al. “Impact of the Ohmic Electrode on the Endurance of Oxide Based Resistive Switching Memory”. In: *IEEE Trans. Electron Devices* 68.3 (2021), pp. 1024–1030.

[214] T. Yonai et al. “98% Endurance Error Reduction by Hard_Verify for 40nm TaOX-based ReRAM”. In: *International Symposium on VLSI Technology, Systems and Application (VLSI-TSA), Hsinchu, TAIWAN*. 2020, pp. 42–43.

[215] K. Maeda et al. “Error recovery of low resistance state in 40nm TaOx-based ReRAM”. In: *International Reliability Physics Symposium (IRPS)*. 2017, 5A-4.1-5A–4.6.

[216] A. Hayakawa et al. “Resolving Endurance and Program Time Trade-Off of 40nm TaOx-Based ReRAM by Co-Optimizing Verify Cycles, Reset Voltage and ECC Strength”. In: *International Memory Workshop*. 2017, pp. 1–4.

[217] M. V. Ganduglia-Pirovano, A. Hofmann, and J. Sauer. “Oxygen vacancies in transition metal and rare earth oxides: Current state of understanding and remaining challenges”. In: *Surf. Sci. Rep.* 62.6 (2007), pp. 219–270.

[218] A. S. Foster et al. “Structure and electrical levels of point defects in monoclinic zirconia”. In: *Phys. Rev. B* 64 (2001), p. 224108.

[219] A.S. Foster et al. “Modelling of point defects in monoclinic zirconia”. In: *J. Non-Cryst. Solids* 303 (2002), pp. 101–107.

[220] W. Kim et al. “3-Bit Multi Level Switching by Deep Reset Phenomenon in Pt/W/TaO_x/Pt-ReRAM Devices”. In: *IEEE Electron Device Lett.* 37.5 (2016), pp. 564–567.

[221] Fu-Cheng Chang et al. “A Study of Read Variability in Backfill Contact Random Access Memory”. In: *International Symposium on VLSI Technology, Systems and Application (VLSI-TSA), Hsinchu, TAIWAN*. 2020, pp. 46–47.

List of Publications

S. Wiefels, U. Böttger, S. Menzel, D. J. Wouters and R. Waser, "Empirical Tunneling Model Describing the Retention of 2.5 Mb HfO_2 based ReRAM", 2020 IEEE International Symposium On VLSI Technology, Systems and Application (VLSI-TSA), Hsinchu, Taiwan 2020.

S. Wiefels, U. Böttger, S. Menzel, D. J. Wouters and R. Waser, "Statistical Modeling and Understanding of HRS Retention in 2.5 Mb HfO2 based ReRAM", 2020 IEEE International Memory Workshop (IMW), Dresden, Germany, 2020, pp. 1-4, doi: 10.1109/IMW48823.2020.9108123.

S. Wiefels, C. Bengel, N. Kopperberg, K. Zhang, R. Waser and S. Menzel, "HRS Instability in Oxide-Based Bipolar Resistive Switching Cells", in IEEE Transactions on Electron Devices, 2020, doi: 10.1109/TED.2020.3018096.

S. Wiefels, M. von Witzleben, M. Hüttemann, U. Böttger, R. Waser and S. Menzel, "Impact of the Ohmic Electrode on the Endurance of Oxide Based Resistive Switching Memory", in IEEE Transactions on Electron Devices, 2021, 10.1109/TED.2021.3049765.

M. Lübben, **S. Wiefels**, R. Waser, I. Valov, "Processes and effects of oxygen and moisture in resistively switching TaOx and HfOx", Advanced Electronic Materials, 2018, 4(1), 1700458, doi: 10.1002/aelm.201700458.

D. Y. Cho, M. Luebben, **S. Wiefels**, K. S. Lee, I. Valov, "Interfacial metal–oxide interactions in resistive switching memories", ACS applied materials & interfaces, 2017, 9(22), 19287-19295, doi: 10.1021/acsami.7b02921.

C. Funck, C. Bäumer, **S. Wiefels**, T. Hennen, R. Waser, S. Hoffmann-Eifert, ... , S. Menzel, "Comprehensive model for the electronic transport in $Pt/SrTiO_3$ analog memristive devices", Physical Review B, 2020, 102(3), 035307, doi: 10.1103/PhysRevB.102.035307.

M. Mayahinia, A. Singh, C. Bengel, **S. Wiefels**, S. Menzel, D. Wouters, ... , S. Hamdioui "A Novel Voltage Controlled Oscillation based ADC Design for Computation-in-Memory Using Emerging ReRAMs". In: Journal on Emerging Technologies in Computing Systems, accepted, 2021.

N. Kopperberg, **S. Wiefels**, S. Menzel, U. Böttger, R. Waser, "Physical Modeling of Instability and Retention Effects in VCM ReRAM Devices Using Kinetic Monte Carlo Methods" (Poster), 9th International RRAM/MRAM Workshop, IMEC, Leuven, 2019